CRIS NLP SERVICE

Library of production-ready applications

UPDATED 10/6/2024

Categorisation updates since last version

1. Trajectory has been added to Outcomes and Clinical Status category

Application updates since last version

Applications	Status
Trajectory	Version 1

APPLICATIONS LIST

INTRODUCTION

This document provides details of natural language processing (NLP) resources which have been developed since around 2009 for use at the South London and Maudsley NHS Foundation Trust (SLaM) NIHR Biomedical Research Centre and its mental healthcare data platform, CRIS.

We have set up the CRIS NLP Service to facilitate the extraction of anonymised information from the free text of the clinical record. Research using data from electronic health records (EHRs) is rapidly increasing and the most valuable information is sometimes only contained in the free text. This is particularly the case in mental healthcare, although not limited to that sector.

CRIS

The Clinical Record Interactive Search (CRIS) system was developed for use within SLaM's NIHR Biomedical Research Centre. It provides authorised researchers with regulated, secure access to anonymised information extracted from SLaM's EHR. SLaM provides mental healthcare to a defined geographic catchment of four south London boroughs (Croydon, Lambeth, Lewisham, Southwark) with around 1.3 million residents, in addition to a range of national specialist services.

Applications to access CRIS and the analyses carried out using CRIS are closely reviewed, monitored and audited by a CRIS Oversight Committee, which carries representation from SLaM's Caldicott Guardian. The CRIS Oversight Committee is responsible for ensuring all research applications comply with ethical and legal guidelines. CRIS was developed with extensive involvement from service users and adheres to strict governance frameworks managed by service users. It has passed a robust ethics approval process acutely attentive to the use of patient data. The data is used in an entirely anonymised and data-secure format and all patients have the choice to opt-out of their anonymised data being used.

CRIS helps us to look at real life situations on a large scale. This means it's easier to see patterns and trends, like what treatments work for some and don't work for others. With this in mind, NLP development has focused particularly on enabling better characterisation of different interventions received (e.g. medications, psychotherapies), the reasons for these interventions (e.g. symptom profiles) and other factors that might affect outcomes (e.g. education, illicit drug use, smoking status).

For more information on CRIS, please have a look at the original or updated protocol papers and the description of its security model and governance framework. Please visit the CRIS website for further information and details of publications.

The CRIS NLP Service

We have developed NLP algorithms (referred to as 'applications' or 'apps' in this document for shorthand) using different approaches, some rules-based and some via machine learning. Other techniques are continually under consideration and evaluation by our own team and in collaboration with teams elsewhere. The General Architecture for Text Engineering (GATE) platform has been used extensively, reflecting a long-running and much-valued collaboration we have had with the University of Sheffield Computer Science Department who originally developed GATE in 1995. Our machine learning algorithms have been greatly facilitated by the TextHunter platform, developed by Richard Jackson, whilst a PhD student at SLaM and KCL, which has allowed annotation at scale for named entity recognition generation.

The purpose of this document is to provide a publicly-accessible and regularly updated resource, containing the details and performance of over 60 NLP applications that we view as 'in production' – i.e. with sufficient description and evaluation to be used across SLaM's and potentially others' EHR data. At any time, a considerable number more are under development and may be cited in publications arising from that development process. Details of these should be sought from authors or the CRIS team.

GENERAL POINTS

All applications currently in production at the CRIS NLP Service are described here. Our aim is to update this document at least twice yearly so please check you are using the version that pertains to the data extraction you are using.

Guidance for use

Applications

Every application report comprises four parts:

1) ***Description*** – the name of application and short explanation of what construct(s) the application seeks to capture.
2) ***Definition*** - an account of how the application was developed (e.g. machine-learning/rule-based, the terms searched for and guidelines for annotators), annotation classes produced and interrater reliability results (Cohen's Kappa).
3) ***Performance*** – precision and recall are used to evaluate application performance in pre-annotated documents identified by the app as well as un-annotated documents retrieved by keyword searching the free text of the events and correspondence sections of CRIS.
   a) Precision is the ratio of the number of relevant (true positive) entities retrieved to the total number of entities (irrelevant -false positive- and relevant -true positive)) retrieved.
   b) Recall is the ratio of the number of relevant (true positive) entities retrieved to the number of relevant (true positive and false negative) entities available in the database.

   Performance testing is outlined in chronological order for either pre-annotated documents, un-annotated documents retrieved through specific keyword searches or both. **The latest performance testing on the list corresponds to results produced by the version of the application currently in use by the NLP Service**. Search terms used for recall testing are presented, where necessary. Similarly, details are provided for any post-processing rules that have been implemented. Notes relating to observations by annotators and performance testers are described, where applicable.
4) ***Production*** – information is provided on the version of the application currently in use by the NLP Service and the corresponding deployment schedule.

Symptom scales (see proposed allocations)
As the number of symptom applications is increasing, we regularly evaluate how to make these available to researchers in a flexible and meaningful manner. To this end, and in order to reduce the risk of too many and/or highly correlated variables in analyses, we are currently utilising symptom scales that group positive schizophreniform, negative schizophreniform, depressive, manic, disorganized and catatonic symptoms respectively. The group of 'other' symptoms represent symptoms that have been developed separately for different purposes and that are intended to be used individually rather than in scales.

Each symptom receives a score of 1 if it's scored as positive within a given surveillance period. Individual symptoms are then summed to generate a total score of:

- 0 – 16 for positive schizophreniform
- 0 – 12 for negative schizophreniform
- 0 – 21 for depressive
- 0 – 8 for manic
- 0 – 8 for disorganized
- 0 – 4 for catatonic

We are encouraging researchers, unless there is a particular reason to be discussed with the NLP team, to use the scales for extracting and analysing data relating to symptom applications.

APPLICATIONS LIBRARY

SYMPTOM SCALES (SEE NOTE)

Positive Schizophreniform	Negative Schizophreniform	Depressive	Manic	Disorganised	Catatonic	Other Symptoms* (Please refer footnote)
Aggression	Anergia	Anergia	Disturbed sleep	Circumstantiality	Echolalia	Anxiety
Agitation	Anhedonia	Anhedonia	Elation	Derailment of speech	Mutism	Bad dreams
Arousal	Apathy	Apathy	Grandiosity	Flight of ideas	Stupor	Cognitive impairment
Delusions	Blunted affect	Disturbed sleep	Insomnia	Formal thought disorder	Waxy flexibility	Hallucinations (visual)
Hallucinations	Concrete thinking	Diurnal variation	Irritability	Loss of coherence		Loneliness
Hallucinations (auditory)	Emotional withdrawal	Early morning wakening	Poor appetite	Poor concentration		Mood instability
Hallucinations (OTG)	Low energy	Guilt	Poor concentration	Tangentiality		Nightmares
Hallucinations (visual)	Negative symptoms	Helplessness	Weight loss	Thought block		Poor insight
Hostility	Poor motivation	Hopelessness				
Irritability	Poverty of speech	Insomnia				
Paranoia	Poverty of thought	Low energy				
Passivity	Social withdrawal	Poor appetite				
Persecutory ideation		Poor concentration				
Thought broadcast		Poor motivation				
Thought insertion		Poverty of speech				
Thought withdrawal		Poverty of thought				
		Social withdrawal				
		Suicidal ideation				
		Tearfulness				
		Weight loss				
		Worthlessness				

FOOTNOTE:
*Other symptoms are intended to be used individually rather than as a scale.

SYMPTOMS

1. AGGRESSION

Description

Application to identify instances of aggressive behaviour in patients, including verbal, physical and sexual aggression.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive mentions include reported to be quite aggressive towards…, violence and aggression, requires continued management and continues to reduce in terms of incidents etc. Also include verbal aggression and physical aggression.

Negative mentions include no aggression, no evidence of aggression etc.

Unknown mentions include unclear statements – aggression won't be tolerated.

Interrater reliability

Cohen's k = 85% (50 un-annotated documents - 25 events/25 attachments, search term 'aggress*')

Search Terms (case insensitive)

aggress

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients, random sample of 50 (one document per patient). 20 documents were evaluated on top of the initial 30 to confirm that precision was low (<80%).	P=73%			
2	Application searches free text for instances of 'aggressi*' only	All patients, random sample of 50 (one document per patient). 20 documents were evaluated on top of the initial 30 to	P=76%			

		confirm that precision was low (<80%).				
3	As above	Random sample of 100 - 15 correspondence-attached text, 4 mental health care plan, 81 event clinical notes	P= 39%	Random sample of 100 - 50 event-clinical note, 50 correspondence-attached text	P=78% R=76%	aggress*
4	As above plus application excludes instances of negation (see notes)	Random sample of 100 - correspondence-attached text, events-clinical notes, risk event description, drug and alcohol history, nurse assessment notes, mental state formulation	P=90%	50 event- clinical note, 50 correspondence-attached text	P=91% R=75%	aggress*

NOTES

Round 3

All false positives in the annotated documents were negations, examples being: ‘no/nil aggression’, ‘no violence or aggression’, ‘no sign of’, ‘did not display/present any’, ‘no arousal, aggression’, ‘no overt aggression’. Other false positives in the non-annotated documents were aggression from others and hearing aggressive voices. Unknowns were comments with a hypothetical ‘may’ or patients having aggressive ideation.

The reason for the higher precision in the non-annotated documents might be because of the documents used. Annotated documents only had 15 correspondence-attached texts while the non-annotated sample used 50. Only two of the false positives in the annotated documents were from correspondence-attached texts. Therefore, false positives (negations of aggression) may be less likely to be picked up in correspondence-attached texts.

The majority of true positives were present mentions of aggression (94.9%) rather than past mentions (eg ‘history of’; 5.1%).

Round 4

Most false positives were due to the negation ‘no’ eg. No violence/aggression or no presentation of violence. Other false positives included aggression that was unrelated to the patient (relative to another patient on the ward), or aggression being in a symptom list (without reference to this being present).

There were not enough false negatives to distinguish a pattern, some instances were: frequent aggressive episodes, risk of aggressive behaviour, was verbally abusive and aggressive.

Code for post-processing

Name like ‘aggressi%’ and *contextstring* not like '%no aggress%' and *contextstring* not like '%nil aggres%'

and *contextstring* not like '%no violence and aggress%'

Production

- Run schedule – monthly
- Version - 1

2. AGITATION

Description

Application to identify instances of agitation.

Definition

Development approach: Machine-learning**.**

Classification of past or present symptom: Both.

Classes produced: Positive, Negative, and Unknown.

Positive mentions include very agitated at present, he was agitated, he was initially calm but then became agitated and started staring and pointing at me towards. Should also include no longer agitated.

Negative mentions include did not seem distracted or agitated, not agitated, no evidence of agitation.

Unknown mentions include unclear statements – a common symptom of psychomotor agitation.

Interrater reliability

Cohen's k = 85% (50 un-annotated documents - 25 events/25 attachments, search term 'agitat*')

Search Terms (case insensitive)

agitat

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients with primary diagnosis code F32* or F33* in a structured field, random sample of 30 (one document per patient)	P=82%			
2		Random sample of 100 - 4 ward progress notes, 11 event-POSProforma, 6 CAMHS event notes, 3 discharge summaries, 22 correspondence-	P=85%	Random sample of 100 - 50 event-clinical note, 50 correspondence-attached text	P=85% R=79%	agitat*

		attached text, 54 events- comments				

NOTES

False positives were mostly when the term 'agitation' was in a list or question with no response of whether the patient experienced it (currently or in the past). Some false positives were negations e.g. 'no episode of...' Psychomotor agitation was classed as unknown. The majority of true positive mentions were present experiences (85.9%) rather than past (14.1%).

Production

- Run schedule – monthly
- Version - 1
- Publications

3. ANERGIA

Description

Application to identify instances of anergia.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive and Negative.

Positive mentions of anergia include feelings of anergia.

Negative mentions of anergia include no anergia, no evidence of anergia, no feeling of anergia.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'anergia*')

Search Terms (case insensitive)

anergia

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients with primary diagnosis code F32* or F33* in a structured field, random sample of 30 (one document per patient)	P=93%			
2		Random sample of 100 - 4 ward progress notes, 2 presenting circumstances, 2 mental state formulation, 2 discharge notification summary, 12 CC correspondence-attached text, 33 correspondence-	P=84%	Random sample of 100 - 51 events- clinical notes, 49 correspondence-attached text	P=95% R=89%	anergia

		attached text, 45 event- clinical note				

NOTES

All false positives occurred due to negations e.g. no loss of interest and anergia, nil anergia, describes no anergia, denies anergia. One unknown was identified as it was vague- unable to assess anergia. The majority of true positives were mentioning anergia as a present symptom (97.6%) rather than a past symptom (2.4%).

Production

- Run schedule – monthly
- Version - 1

4. ANHEDONIA

Description

Application to identify instances of anhedonia (inability to experience pleasure from activities usually found enjoyable).

Definition

Development approach: Machine-learning**.**

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive mentions of anhedonia or anhedonic symptoms include X had been anhedonic, X has anhedonia.

Negative mentions of anhedonia or anhedonic symptoms include no anhedonia, no evidence of anhedonia, not anhedonic.

'Unknown' annotations included: i) used in a list, not applying to patient (e.g. typical symptoms include …); ii) uncertain (might have anhedonia, ?anhedonia, possible anhedonia); iii) not clearly present (monitor for anhedonia, anhedonia has improved); iv) listed as potential treatment side-effect; v) vague ('she is not completely anhedonic', 'appears almost anhedonic')

Interrater reliability

Cohen's k=85% (50 un-annotated documents - 25 events/25 attachments, search term 'anhedon*')

Search Terms (case insensitive)

anhedon

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients with primary diagnosis code F32* or F33* in a structured field, random sample of 30 (one document per patient)	P=87%			
2		Random sample of 100 - 4 ward progress notes, 1 presenting circumstances, 1 mental health care plan, 16 CCS	P=94%	Random sample of 100 - 50 events- clinical notes, 50 correspondence-attached text	P=93% R=86%	anhedon*

		correspondence-attached text, 36 correspondence-attached text, 42 events- clinical note				

NOTES

The majority of false positives occurred when the negation 'nil' was used, sometimes when the term 'denies' was used also. Unknown was classified when mentioning 'partial' anhedonia due to a chronic illness. All positives were current symptoms rather than past tense (history of anhedonia).

Production

- Run schedule – monthly
- Version - 1

5. ANOSMIA

Description

Application to extract and classify mentions related to anosmia.

Definition

Development approach: machine learning***.***

Classification of past or present symptom: Both.

Classes produced: positive, negative, unknown, form.

Positive examples: Annotations are coded as positive when there is a reference to symptoms/experiences of anosmia

E.g.

Loss of enjoyment of food due to anosmia

COVID symptoms such as anosmia

Negative examples: Annotations are coded as negative when there is no reference to symptoms/experiences of anosmia

E.g.

Nil anosmia

Doctor mentioned they had anosmia so could not smell patient

Anosmia related to people other than the patient

Unknown examples: Annotations are coded as unknown when it is not clear if the patient has symptoms/experiences of anosmia

E.g.

Mentions of medications for it

Form examples: Annotations are coded as form when there is reference of anosmia in terms of an automated letter or email between colleagues or

E.g.

Don’t come to the practice if you have any covid symptoms such as anosmia, etc.

Interrater reliability

N/A only one annotator

Search Terms (case insensitive)

Anosmia*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 patients	P=85%	Random sample of 100 events and attachments	P=83% R=93%	Anosmia*

NOTES

Production

- Run schedule – on demand
- Version – 1

6. ANXIETY

Description

Application to extract and classify mentions related to (any kind of) anxiety.

Definition

Development approach: Rule-based***.***

Classification of past or present symptom: Both.

Classes produced: Affirmed, negated and irrelevant.

a. affirmed/negated/irrelevant ("status"),
   i. affirmed: "ZZZ shows **anxiety** problems"
   ii. negated: "ZZZ does not show **anxiety** problems"
   iii. irrelevant: " **if** ZZZ was **anxious** he would not take his medication"
b. related to the patient or someone else ("experiencer")
   i. patient: "ZZZ shows **anxiety** problems"
   ii. other: "nurse is **worried** about the patient"
   iii. unknown: "he showed clear signs of **anxiety"**
c. an objective or subjective mention ("observation")
   i. objective:"ZZZ showed signs of anxiety today"
   ii. subjective: "ZZZ says he feels anxious"

Detailed annotation guidelines with further examples are available on request.

The NLP algorithm is trained only to classify affirmed/patient as 1, rest as 0. Although this is not what the annotations reflect, the annotations could be used to train new algorithms for the other attributes.

Interrater reliability

3000 annotations

- **status** 83 kappa, 94 IAA
- **experiencer** 81 kappa, 93 IAA
- **observation** 61 kappa, 76 IAA

Search Terms (case insensitive)

Available on request

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 - 1 Addictions event, 110 attachments, 14 CAMHS events, 3 mental health care	P=94%	Random sample of 100 events and attachments	P=87% R=97%	anxious anxiet* worried restless

		plans, 6 CCS_correspondance, 1 discharge notification summary, 95 events, 6 history, 4 mental health formulations, 3 summaries of need, 33 ward progress notes)				worry

NOTES

False positives - - negations with the word 'not' e.g. 'not being anxious', 'did not experience worry', 'Mood, Anxiety and Personality Disorder Clinical Academic Group' mention (most consistent FP). Also, instances relating to relatives being worried/anxious.

Production

- Run schedule – monthly
- Version – 1

7. APATHY

Description

Application to extract the presence of apathy.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive mentions include any indication that apathy was being reported as a symptom: e.g. continues to demonstrate apathy; symptoms include apathy he is withdrawn, attributable to apathy; his apathy … ; some degree of apathy noted; presentation with apathy; his report of apathy given patient's level of apathy. Most apathy statements were found to be accompanied by 'negative symptoms' (i.e. rather than depressive). Should include implicit mentions of previous apathy, if evaluating on past or present.

Very few negative mentions found. Usual statements (denied apathy; no evidence of apathy etc.)

'Unknown' annotations include apathy mentioned as a hypothetical cause of something else (e.g. inactivity) rather than described as being present; apathy mentioned as a possibility in the future (e.g. may develop A apathy or as a possible side effect of medication (rather than actually present), or as an early warning sign. Also *apathy* found in quite a few names.

Interrater reliability

Cohen's k=86% (50 un-annotated documents - 25 events/25 attachments, search term 'apath*')

Search Terms (case insensitive)

apath

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients with primary diagnosis code F32* or F33* in a structured field, random sample of 30 (one document per patient)	P=70%			
2	Application searches free text for instances of 'apathy' or	All patients with primary diagnosis code F32* or F33* in a structured field, random	P=73%			

	'apathetic' only (see notes)	sample of 30 (one document per patient)				
3	As above	Random sample of 100 - 4 ward progress notes, 1 presenting circumstances, 1 mental health care plan, 16 CCS correspondence-attached text, 36 correspondence-attached text, 42 events- clinical note	P=94%	Random sample of 100 - 50 events- clinical notes, 50 correspondence-attached text	P=93% R=86%	apath*

NOTES

False positives occurred when the negation 'denies' apathy came up. Unknowns were classified when the vague 'maybe' term was used or the symptom was present in a list without response on whether the symptom was present or not. Most true positives were current symptoms (99%) rather than past tense (history of apathy).

Code for post-processing

Name like 'apathy' or *name* like 'apathetic'

Production

- Run schedule – monthly
- Version – 1

8. AROUSAL

Description

Application to identify instances of arousal excluding sexual arousal.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive mentions include physiological, emotional and autonomic hyperarousal such as "...the decisions she makes when emotionally aroused", "...during hyperaroused state", "following an incidence of physiological arousal"

Negative mentions include mentions of sexual arousal, no arousal, not aroused, denies being aroused, less aroused, less arousal, low arousal.

Unknown mentions: annotations include unclear statements and hypotheticals ("if aroused...")

Interrater reliability

Cohen's k = 95% (50 un-annotated documents - 25 events/25 attachments, search term '*arous*')

Search Terms (case insensitive)

arous

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		Random sample of 100 – CAMHS events	P=71%	Random sample of 100 - 50 events- clinical notes, 50 correspondence-attached text	P=89% R=91%	*arous*

NOTES

False positives mainly occurred when referencing sexual arousal or negation (did not arouse, no symptom of…, low arousal, unarousable). Other false positives related to arousal of someone other than the patient. Unknowns included hyper-arousal to specific stimuli e.g. due to PTSD diagnosis, hypothetical mention, arousal included in list (without direction if hypo/hyper arousal), arousal scores or description of arousal task administered without comment on the outcome.

Production

- Run schedule – monthly
- Version - 1

9. BAD DREAMS

Description

Application to identify instances of experiencing a bad dream.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include any mention of the patient having a nightmare or bad dream e.g. 'ZZZZZ had a bad dream last night', 'she frequently has bad dreams', 'ZZZZZ has suffered from bad dreams in the past', 'ZZZZZ had a bad dream that she was underwater', ' he's been having fewer bad dreams'

Negative annotations include instances where a bad dream has not occurred, metaphorical comparisons: 'she denied any bad dreams', 'does not suffer from bad dreams', 'no other PTSD symptoms such as bad dreams', 'he said the experience was like a bad dream', 'ZZZZZ compared his time in hospital to a bad dream', 'said she wanted to wake up from this bad dream'

Unknown annotations include instances where it is not clear whether a bad dream has occurred: 'she said it might have been a bad dream', 'he woke up in a start, as if waking from a bad dream', 'ZZZZZ couldn't remember whether the conversation was just a bad dream', 'doesn't want to have bad dreams'

Interrater reliability

Cohen's k = 100% (100 unannotated documents- 50 events/50 attachments, search terms 'dream' and 'dreams')

Search Terms (case insensitive)

bad dream*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
2		Random sample of 100 – CAMHS event-comments	P=92%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P=89% R=100%	dream dreams

Production

- Run schedule – monthly
- Version - 1

10. BLUNTED AFFECT

Description

Application to identify instances of blunted affect.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive and Negative.

Positive annotations include his affect remains very blunted, objectively flattened affect, states that ZZZZZ continues to appear flat in affect. Include affect somewhat flat.

Negative annotations include incongruent affect, stable affect, no blunted affect.

Unknown annotations include 'typical symptoms include blunted affect', 'slightly flat affect', 'relative shows flat affect'.

Interrater reliability

Cohen's k = 100% (50 annotated documents - 25 events/24 attachments/1 mental health care plan)

Search Terms (case insensitive)

affect

Blunt* [0 to 2 words in between] *affect*

Flat [0 to 2 words in between] *affect*

Restrict [0 to 2 words in between *affect*

affect [0 to 2 words in between] blunt

affect [0 to 2 words in between] flat

Affect [0 to 2 words in between] restrict

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients, random sample of 30 (one document per patient)	P=93%			
2		Random sample of 100 - 25 ward progress notes, 4 assessment-	P=98%	Random sample of 100 - 50 events- clinical notes, 50	P=100% R=80%	affect

		mental state comments, 1 mental health care plan, 22 correspondence-attached text, 48 events- clinical note		correspondence-attached text		

NOTES

The few false positives seen were irrelevant mentions of ‘flat’ in relation to the patients’ living situation (that is ‘affecting’ them). All true positives reflected current presentation rather than past (history of blunted affect) I the annotated documents.

Production

- Run schedule – monthly
- Version - 1

11. CIRCUMSTANTIALITY

Description

Application to identify instances of circumstantiality.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive mentions include signs of over inclusiveness and circumstantially, loose associations and circumstantiality, circumstantial in nature. Also include some circumstantiality at points and speech is less circumstantial.

Negative mentions include no signs of circumstantiality, no evidence of circumstantial.

Unknown mentions include circumstantial mentioned as a hypothetical cause of something else.

Interrater reliability

Cohen's k = 100% (50 annotated documents - 25 events/25 attachments)

Search Terms (Case insensitive)

circumstan

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients, random sample of 30 (one document per patient). 20 documents were evaluated on top of the initial 30 per evaluation to confirm that	P=38%			

		precision was low (<80%).				
2	Application excludes instances of 'circumstance*' (see notes)	All patients, random sample of 30 (one document per patient). 20 documents were evaluated on top of the initial 30 per evaluation to confirm that precision was low (<80%).	P=90%			
3	As above	Random sample of 100 - 9 ward progress notes-comments, 5 mental state formulation, 1 CCS corresponden ce- attached text, 28 corresponden ce- attached text, 57 events-comments	P=97%	Random sample of 100 - 50 events-clinical notes, 50 corresponde nce-attached text	P=94% R=92%	circumstant*

NOTES

False positives occurred with certain negations e.g. no circumstantiality, and with irrelevant mentions e.g. circumstantial evidence. All positive mentions were current instances of circumstantial speech. False negatives were mentions of circumstantial thought.

Code for post-processing

Name not like 'circumstance%'

Production

- Run schedule – monthly
- Version - 1

12. COGNITIVE IMPAIRMENT

Description

Application to identify instances of cognitive impairment. The application allows to detect cognitive impairments related to attention, memory, executive functions, and emotion, as well as a generic cognition domain. This application has been developed for patients diagnosed with schizophrenia.

Definition

Development approach: Rule-based.

Classification of past or present symptom: Both.

Classes produced: affirmed and relating to the patient and negated/irrelevant.

Below are specific examples for the attention domain:

key	score	comment
patient shows attention problems (positive)	1	affirmed
patient does not shows attention problems or shows good attention (negated)	-1	negated
irrelevant	0	irrelevant

example	score	comment
ZZZ shows good concentration	-1	
ZZZ does not show good concentration	1	
ZZZ shows poor concentration	1	
patient scored 10/18 for attention	1	anything below maximum score classified as 1
ZZZ seems distracted at times/absent-minded	1	
patient received diagnosis of adhd	1	
ZZZ seems more distracted than yesterday	1	worsenings noted as 1
ZZZ seems less distracted today	1	improvement noted as 1
ZZZ's concentration has improved since last session	1	improvement noted as 1
ZZZ said the voices distract him	1	poor concentration caused by hallucinations noted as 1
ZZZ seems hyper-vigilant	1	
the session focused on…	0	examples of frequent irrelevant uses
patient uses distraction technique to ignore hallucinations	0	
concentration camp	0	
attention seeking	0	
pay attention to	0	
patients needs (medical) attention	0	
bring to someone's attention	0	
draw your attention to…	0	
keywords: attention, concentration, focus, distracted, adhd, hypervigilance, attend to		

Interrater reliability

Cohen's k

Cognition - 66% (3000 un-annotated documents)

Emotion – 84% (3000 un-annotated documents)

Executive function – 40% (3000 un-annotated documents)

Memory – 68% (3000 un-annotated documents)

Attention – 99% (2616 un-annotated documents)

Search Terms (Case insensitive)

Gazetteer available on request

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 100 - 42 attached text, 4 CAMHS events, 1 CCS correspondence, 1 referral reason, 11 ward progress notes, 1 summary of need, 1 mental health examination, 1 risk assessment tool, 38 event comments	P=80%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P=96% R=92%	prompt* manag* memory *understand* cogniti*
2		Patients with an F20 diagnosis sample of 100 - 36 event attachments, 3 CCS correspondence, 1 brief notification summary, 5 risk	P=84%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P=78% R=70%	prompt* manag* memory *understand* cogniti*

		assessment tool, 2 summary of needs, 1 ward round, 22 ward progress notes, 20 event comments				

NOTES

Lower precision and recall in the F20 sample most likely due to the low number of true positives present in the sample.

Production

- Run schedule – on request
- Version - 1

13. CONCRETE THINKING

Description

Application to identify instances of concrete thinking.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include text referring to 'concrete thinking', speech or answers to questions being 'concrete', the patient being described as 'concrete' without elaboration, answers being described as concrete in cognitive assessments, 'understanding' or 'manner' or 'interpretations' of circumstances being described as concrete. This included episodes in the past and both strong and weak (e.g. 'tendency to concrete interpretations') manifestations.

Negative annotations include 'no evidence of concrete thinking'

Unknown annotations include references to concrete as a material (concrete floor, concrete house etc.), 'no concrete plans' referring to suicidal ideation, delusions being concrete, 'no concrete symptomatology', achieving 'concrete goals', using 'concrete learning activities'.

Initially, we used the keyword 'concrete*' to pick up instances of concrete thinking. But when we manually completed the first round of annotations, performance was not satisfactory. After checking positive and negative annotations, some regular patterns were identified whereby the word 'concrete' was used within one or two words before or after the word 'thinking' which informed the final choice of search terms below.

Interrater reliability

Cohen's k = 83% (50 un-annotated documents - 25 events/25 attachments, search term 'concrete*')

Search Terms (Case insensitive)

Concrete [word][word]think*

think [word] [word] concret*

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients, random sample of 30 (one document per patient)	P=87%			

2		Random sample of 146 - 57 correspondence-attachment text, 14 CAMHS event-comments, 38 events-comments, 36 care plan-outcome detail (mental health)	P=91%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P=84% R=41%	concrete

NOTES

False positives occurred when statements were irrelevant, such as concrete thinking of a relative, a rehabilitation plan or therapeutic task. The term 'no evidence of' was also ignored when relating to concrete thinking. Undetected comments include mentions of a patient being 'rigid and concrete', 'socially concrete', 'rigid in way of answering questions, 'concrete in thought' and 'concrete in vocabulary use'. Comments were annotated as unknown when they were hypothetical 'may have concrete thinking' or described as 'sometimes' concrete, without specifying whether they generally the patient generally is or not.

Production

- Run schedule – monthly
- Version - 1

14. DELUSIONS

Description

Application to identify instances of delusions.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive mentions include paranoid delusions; continued to express delusional ideas of the nature etc. Also include no longer delusional- indicates past.

Negative mentions include no delusions, denied delusions.

Unknown mentions include delusions are common.

Interrater reliability

Cohen's k = 92% (50 un-annotated documents - 25 events/25 attachments, search term 'delusion*')

Search Terms (case insensitive)

delusion

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 30 (one document per patient)	P=87%			
2		Random sample of 100 - 22 ward progress notes, 1 discharge summary, 26 correspondence- attached text, 49 event-clinical note	P=97%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P=77% R=86%	delusion*

3	Application excludes instances of negation – '*no delusion*', '*not expressed any delusion*', '*didn't express any delusion*' (see notes)	Random sample of 100 – 26 ward progress note, 1 mental state formulation, 2 discharge notification summaries, 1 mental health care plan, 40 correspondence- attached text, 30 events-clinical note	P=90%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P=93% R=85%	delusion*

NOTES

False positives occurred due to negations e.g. not seen to be, no evidence of, not expressed, no…or delusions, no overt delusional behaviour. Other false positives were unclear mentions e.g. when symptoms were in a list, possibly…, understanding if there Is presence of… Ignoring the 'seem to be' and 'expressed' mentions there was not enough consistency in false positives to decipher a pattern. There were only four false positives, two involving the word 'expressed'. The other two were 'appeared quite delusional' and 'delusional sexual themes have diminished'.

Code for post-processing

contextstring not like '%no delusion%' and *contextstring* not like '%not expressed any delusion%' and *contextstring* not like '%didn't express any delusion%'

Production

- Run schedule – monthly
- Version - 1

15. DERAILMENT

Description

Application to identify instances of derailment.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include he derailed frequently, there was evidence of flight of ideas and thought derailment in his language etc. Include past mentions e.g. 'speech no longer derailed'.

Negative annotations include the thought stream is normal as he uses sentences in consequences with no derailment, erratic compliance can further derail her stability etc. Also include no evidence of derailment, without derailment, without derailing, no derailment, no thought block, derailment, tangentiality noted, no evidence of loosening of association, derailment or tangential thoughts.

Unknown annotations include train was derailed.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'derail*')

Search Terms (case insensitive)

derail

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1	Application excludes derailing of trains, trams and efforts to achieve goals	All patients, random sample of 30 (one document per patient)	P=74%			
2	As above	Random sample of 100 – 1 assessment note, 8 risk event descriptions,	P=73%	Random sample of 100 - 50 events- clinical notes, 50 corresponde	P=88% R=95%	derail*

		9 ward progress notes, 1 CCS corresponden t- attached text, 1 discharge notification summary, 3 CAMHS event- clinical note, 35 corresponden ce- attached text, 29 events- clinical notes		nce- attached text		
3	As above	Random sample of 100 – 6 discharge notification summaries, 3 mental state comments, 1 nurse assessment notes, 26 corresponden ce-attached text, 64 event-clinical note	P=87		P=84% R=99%	derail*

NOTES

False positives mainly occurred due to negations e.g. 'no evidence of', 'no sign of' or simply 'no derailment'. False positives also occurred when mentions were irrelevant e.g. derail treatment, derail a session, another individual derailing a session. Unknown was labelled for one unsure mention of a vague term use; 'I wonder'. The majority of true positives was of derailment being a current symptom. Precision was high in non-annotated documents, as there were only 11 negatives. However, they were all flagged as positive. This is probably due to the app not computing negations. Regarding recall, positives were not flagged in mentions where derailment was at the beginning of a short sentence e.g. 'Derailment.'.

Production

- Run schedule – monthly
- Version - 1

16. DISTURBED SLEEP

Description

Application to identify instances of disturbed sleep.

Definition

Development approach: Rule-based***.***

The application identifies instances of disturbed sleep as follows: complains of poor sleep, poor sleep, sleep disturbed, sleep difficulty, sleeping poorly, not sleeping very well, cannot sleep, sleep pattern poor, difficulties with sleep, slept badly last couple of nights.

Interrater reliability

Cohen's k = 75% (50 un-annotated documents - 25 events/25 attachments, search term '*sleep*' or 'slept')

Search Terms

Not
poor*
interrupt*
disturb*
inadequat*
disorder*
prevent*
stop*
problem*
difficult*
reduc*
less*
impair*
erratic*
unable*
worse*
depriv*
[0-2 token]
sleep* or slep*
little sleep
sleepless night
broken sleep
sleep intermittently
sleep* or slep*

[0-2 token] not
poor*
interrupt*
disturb*
inadequat*
disorder*
prevent*
stop*

problem*
difficult*
reduc*
less*
impair*
erratic*
unable*
worse*
depriv*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 – 2 mental state formulation, 1 CCs correspondence-body text, 3 discharge summaries, 1 mental health care plan, 1 presenting circumstances, 1 risk event, 2 brief summaries, 36 correspondence-attached text, 53 events	P=89%	Random sample of 100 - 50 events- clinical notes, 50 correspondence-attached text	P=88% R=68%	*sleep* slept

NOTES

False positives included negation (denies, no…sleep disturbance, …not disturbing sleep), sleeping tablets (extra dose to sleep, taking tables not to sleep but other intention), hypotheticals e.g. risk of poor sleep. No pattern observed in false negatives. Examples include sleep - reported as disturbed, reported sleeping only 4 hours a night, he didn't sleep through the night, his sleep has deteriorated.

Production

- Run schedule – monthly
- Version - 2

17. DIURNAL VARIATION OF MOOD

Description

Application to identify instances of diurnal variation of mood.

Definition

Development approach: Machine-learning**.**

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive examples:

- 'patient complaints of diurnal variation'
- 'he reported diurnal variation in his mood'
- 'Diurnal variation present'
- 'some diurnal variation of mood present'

Negative examples:

- 'no diurnal variation'
- 'diurnal variation absent'
- 'patient complaints of ongoing depression but no diurnal variation'
- 'depressive symptoms present without diurnal variation'

Unknown examples:

- 'diurnal variation could be a symptom of more severe depression'
- 'we spoke about possible diurnal variation in his mood'
- 'it was not certain if there were diurnal variation'

Interrater reliability

N/A

Search Terms (case insensitive)

diurnal variation

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 documents – 44 attachments, 8 CCS correspondence, 1 discharge notification, 40 events, 2 mental state formulation, 1 presenting	P=86%	Random sample of 100 events and attachments	P=94% R=100%	diurnal

		circumstances, 2 single generic assessments, 2 progress ward notes				

Production

- Run schedule – monthly
- Version - 1

18. DROWSINESS

Description

Application to identify instances of drowsiness.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive examples:

- ZZZZZ appeared to be drowsy.
- She has complained of feeling drowsy.
- Positive annotations should be anything implying that the patient is or has been drowsy or showing/reporting drowsiness. The timing doesn't matter (i.e. past or present).

Negative examples:

- He is not drowsy in the mornings.
- She was quite happy and did not appear drowsy.
- ZZZZZ denied any symptoms of drowsiness.
- Negative annotations should be when the patient denies drowsiness, or is described as not drowsy etc.

Unknown examples:

- In reading the label (of the medication), ZZZZZ expressed concern in the indication that it might make him drowsy
- Monitor for increased drowsiness and inform for change in presentation.
- The 'unknown' category of annotations should be everything else. This would include hypothetical statements (e.g. risk of drowsiness, instructions to reduce medication if the patient becomes drowsy) or anything else insufficiently certain (this includes statements like 'he is less drowsy now' – although this implies that the patient was once drowsy, this isn't really clear enough [although 'he is more drowsy' does imply drowsiness]). Also, there may be mentions about drowsiness in people other than the patient (e.g. relatives).

Interrater reliability

Cohen's k = 83% 1000 un-annotated documents, search term 'drows*')

Search Terms

drows*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 documents	P=91%	Random sample of 100 documents	P=77% R=93%	drows*

2	The application excludes ‘no drows*’, ‘wasn’t drows*’, ‘no reported drows*’			Random sample of 100 documents	P= 80% R=100%	drows*

Production

- Run schedule – monthly
- Version - 1

19. EARLY MORNING WAKENING

Description

Application to identify instances of early morning wakening.

Definition

Development approach: Machine-learning**.**

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations:

- 'patient complaints of early morning awakening'
- 'he reported early morning wakening'
- 'Early morning awakening present'
- 'there is still some early morning wakening'

Negative annotations:

- 'no early morning wakening'
- 'early morning wakening absent'
- 'patient complaints of disturbed sleep but no early morning awakening'
- 'sleeps badly but without early morning wakening

Unknown annotations:

- 'early morning awakening could be a symptom of more severe depression'
- 'we spoke about how to deal possible early morning wakening'
- 'it was not certain if there were occasions of early morning awakening'

Interrater reliability

N/A

Search Terms (case insensitive)

early morning wakening

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 – 36 attachments, 2 care plans mental health,	P=96%	Random sample of 100 events and attachments	P=95% R=96%	*wakening

		1 CAMHS event, 10 CCS correspondence, 2 discharge notifications summaries, 41 events, 1 presenting circumstance, 1 single generic assessment, 4 summaries of need, 1 triage form ARC, 1 ward progress note				

Production

- Run schedule – monthly
- Version - 1

20. ECHOLALIA

Description

Application to extract occurrences where echolalia is present.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include no neologisms, but repeated what I said almost like echolalia, intermittent echolalia. Also include some or less echolalia.

Negative annotations include no echolalia, no evidence of echolalia, cannot remember any echolalia or stereotyped utterances.

Unknown annotations include echolalia is not a common symptom. Also include hypotheticals such as he may have some echolalia, evidence of possible echolalia.

Interrater reliability

Cohen's k = 88% (50 un-annotated documents - 25 events/25 attachments, search term 'echola*')

Search Terms (case insensitive)

echola

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1	Application searches free text for instances of 'echolali*' (see notes)	All patients, random sample of 30 (one document per patient). 20 documents were evaluated on top of the initial 30 to confirm that precision was low (<80%).	P=74%			
2	As above	Random sample of 100 –	P=96%	Random sample of	P=89%	echola*

		18 ward progress note, 1 discharge notification summary, 38 correspondenc e- attached text, 19 CCS correspondenc e- attached text, 24 events-clinical note		100 - 50 events-clinical notes, 50 corresponde nce-attached text	R=86%	

NOTES

False positives occurred with certain negations e.g. does not demonstrate/display. Unknowns were echolalic pathological laughter and when echolalia was questioned e.g. could be echolalia, echolalia? False negatives were a suggestion of echolalia, uses echoed speech, Echolalia (one-word statement), regularly echoed words. The majority of true positives in the annotated document was present tense, only 1% past echolalia.

Code for post-processing

Name like 'echolali%'

Production

- Run schedule – monthly
- Version - 1

21. ELATION

Description

Application to identify instances of elation.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include mildly elated in mood, elated in mood on return from leave, she appeared elated and aroused etc.

Negative annotations include ZZZZZ was coherent and more optimistic/aspirational than elated throughout the conversation, no elated behaviour etc.

Unknown annotations include unclear statements such as in his elated state there is a risk of accidental harm, 'monitor for elation'. Should also include statements listed as potential treatment side-effects 'elation is a known side effect' and statements were term is used in a list, not applying to patients (e.g. Typical symptoms include...).

Interrater reliability

Cohen's k = 100% (50 annotated documents - 25 events/25 attachments)

Search Terms (Case insensitive)

elat

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1	Application searches free text for instances of 'elated*' or 'elation*' (see notes)	All patients, random sample of 30 (one document per patient)	P=90%			
2	As above	Random sample of 100 – 5 ward progress notes, 1 presenting	P=95%	Random sample of 100 - 50 events-clinical	P=94% R=97%	elat*

		circumstance mention, 1 CCS correspondence- attached text, 1 mental health care plan, 23 correspondence- attached text, 69 events-comments		notes, 50 correspondence-attached text		

NOTES

False positives occurred for two negations: ‘nor elation’ and ‘not elated’. Unknowns were classed for mentions stating ‘monitor for elation’ and ‘if any evidence of elated mood’. False negatives: Mentions not flagged by app as positive: ‘was elated’, ‘get elated’, and ‘elated mood’. However, each of these only occurred once. The majority of true positive were current mentions of elation (98%) rather than past.

Code for post-processing

name like 'elated%' or 'elation%'

Production

- Run schedule – monthly
- Version - 1

22. EMOTIONAL WITHDRAWAL

Description

Application to identify instances of emotional withdrawal.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations apply to any description of the patient being described as withdrawn or showing withdrawal but with the following exceptions (which are annotated as unknown):

- Alcohol, substance, medication withdrawal
- Withdrawal symptoms, fits, seizures etc.
- Social withdrawal (i.e. a patient described as becoming withdrawn would be positive but a patient described as showing 'social withdrawal' would be unknown – because social withdrawal is covered in another application).
- Thought withdrawal (e.g. 'no thought insertion, withdrawal or broadcast')
- Withdrawing money, benefits being withdrawn etc.

Negative and unknown annotations are restricted to instances where the patient is being described as not withdrawn and categorised as unknown.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'withdrawn')

Search Terms (case insensitive)

withdrawn

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 – 2 CAMHS events-comments, 2 discharge notifications, 2 mental health care plans, 9	P=87%	Random sample of 100 - 50 events-clinical notes, 50 corresponde nce-	P=85%, R=96%	withdrawn

		ward progress notes, 24 correspondence- attached text, 61 correspondence- attached text		attached text		

NOTES

False positives were related to irrelevant mentions e.g. Police statement withdrawn, money withdrawn, specific named drug withdrawn, appointment withdrawn, contact withdrawn. However, this did not occur many times.

Production

- Run schedule – monthly
- Version - 1

23. EYE CONTACT (CATEGORISATION)

Description

Application to identify instances of eye contact and determine the type of eye contact

Definition

Development approach: Rule-based***.***

Classification of past or present symptom: Both.

Classes produced: Positive and Negative.

Positive mentions: the application successfully identifies the type of eye contact (as denoted by the keyword) in the context (as denoted by the contextstring)
e.g., keyword: 'good'; contextstring: 'There was good eye contact'

Negative mentions: the application does not successfully identifies the type of contact (as denoted by the keyword) in the context (as denoted by the contextstring). The keyword does not related to the eye contact
e.g., keyword: 'showed'; contextstring: 'showed little eye-contact'.

Interrater reliability

Not applicable

Search Terms (case insensitive)

Eye *contact*

Keyword: the term describing the type of eye contact

ContextString: the context containing the keyword in its relation to eye-contact

Performance

	Post-processing rules added to application	Annotated documents	Performance (annotated)	Un-annotated documents from keyword search in CRIS	Performance (un-annotated)	Keywords used for random extraction from CRIS
1		All patients, random sample of 100 (one document per patient – 25 attachments, 4 CAMHS events, 1 CCS correspondence, 6 discharge notification summary, 45 events, 1 mental state formulation, 2	P=86%	Random sample of 100 – 50 attachments, 50 events	P=91% R=80%	eye *contact*

		single generic assessments, 1 summary of need, 13 progress ward notes, 2 ward rounds				

NOTES

Not applicable

Production

- Run schedule – on request
- Version - 1

24. FATIGUE

Description

Application to identify symptoms of fatigue.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include positive mentions of fatigue symptoms/experiences when it is clearly defined that symptom is experienced by patient. It will not consider fatigue as positive if it is given as side effect of medication or any other symptom**.**

Positive examples: 'zzz has been experiencing fatigue'; 'fatigue interfering with daily activities'.

Negative annotations include situations when there is no reference to symptoms/experiences of fatigue, fatigue related to people other than the patient, or fatigue as side effects of mediations/treatment

Negative examples: 'no mentions of fatigue'; 'her high levels of anxiety impact on fatigue'; 'main symptoms of dissociation leading to fatigue'

Unknown annotations include situations when it is not clear if the patient has symptoms/experiences for fatigue.

Unknown examples: 'zzz is undertaking CBT for fatigue'.

Interrater reliability

Not applicable

Search Terms (case insensitive)

Fatigue, exclude 'chronic fatigue syndrome'

Performance

	Post-processing rules added to application	**Annotated documents**	**Performance (annotated)**	**Un-annotated documents from keyword search in CRIS**	**Performance (un-annotated)**	**Keywords used for random extraction from CRIS**
1		All patients, random sample of 100 – 48 Attachment_Text, 44 Comments, 1 Description, 1 Notes, 1 Reason_For_Referral_And_Biographical_Information, 1	P=79%	Random sample of 100 – 50 attachments, 50 events	P=78% R=95%	Fatigue*

		risk_event_description, 2 Summary_Of_Need				

NOTES

Not applicable

Production

- Run schedule – montly

25. FLIGHT OF IDEAS

Description

Application to extract instances of flight of ideas.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include Mrs ZZZZZ was very elated with by marked flights of ideas; marked pressure of speech associated with flights of ideas. Also include 'some flight of ideas'.

Negative annotations include no evidence of flight of ideas, no flight of ideas.

Unknown annotations include 'bordering on flight of ideas', or when used in a list not applying to the patient 'typical symptoms include', or irrelevant mentions 'relative shows FOI'.

Interrater reliability

Cohen's k = 96% (50 un-annotated documents - 25 events/25 attachments, search term 'flight of')

Search Terms (case insensitive)

flight *of* *idea*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 50 (one document per patient). 20 documents were evaluated on top of the initial 30 to confirm that precision was low (<80%).	P=72%			
2		Random sample of 100 – 9 ward	P=89%	Random sample of 100 - 50	P=91%, R=94%	flight of

		progress notes, 1 risk event description, 5 mental health care plans, 23 correspondenc e- attached text, 62 event- clinical notes		events- clinical notes, 50 corresponde nce- attached text		

NOTES

False positives occurred when negations were mentioned e.g. no obvious flight of ideas, no flight of ideas, no evidence of …. or flight of ideas. Unknowns occurred when the symptom was mentioned in a list without reference to it being present and when it was being questioned. The majority of true positives were present tense mentions (95% in annotated documents). There were only three instances where the app did not flag a mention as positive (high recall).

Production

- Run schedule – monthly
- Version - 1

26. FLUCTUATION

Description

The purpose of this application is to determine if a mention of fluctuation within the text is relevant

i.e., associated with the patient and affirmed.

Example: '… patient's mood has been fluctuating…'

No particular diagnosis group was chosen for the development of this application

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, negative, and unknown.

Positive annotations include 'Mrs ZZZZZ's mood has been fluctuating a lot', or 'suicidal thoughts appear to fluctuate'

Negative annotations include 'no evidence of mood fluctuation' or 'does not appear to have significant fluctuations in mental state'

Unkown annotations include 'unsure whether fluctuation has a mood component' or 'monitoring to see if fluctuations deteriorate' or 'his mother's responsibility fluctuated' or 'is the person's risk likely to fluctuate.. yes/no…'

Interrater reliability

No. of annotations per annotator:

Annotator	Attachments	Events
Annotator 1	1387	1113
Annotator 2	2751	2243
Annotator 3	1862	458
Total	6000	3814

Not all were double annotated. Out of these, 4,402 (pre-adjudication) annotations (doubly annotated by medical students) were used.

No of annotations in attachments after adjudication - 3,140

Inter-annotator agreement metrics:

Document type	class					subject				
	p	r	f1	acc	k	p	r	f1	acc	k

Attachment	0.93	0.91	0.92	0.91	0.85	0.86	0.88	0.86	0.88	0.87
Attachment2	0.94	0.94	0.94	0.94	0.89	0.94	0.93	0.93	0.93	0.92
Event	0.96	0.95	0.96	0.95	0.91	0.95	0.95	0.95	0.95	0.94

Search Terms (case insensitive)

fluctuat

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 100 – 39 attachment text, 2 CAMHS Event, 2 CCS_correspondence, 1 Discharge_Notification_Summary, 38 Event, 1 Mental_state_formulation, 3 Single_generic_Assessment, 1 Summary_Of_Need, 13 Ward Progress Note	P=96%	Random sample of 100 – 50 Event, 50 Attachment	P=87% R=96%	Fluctuation*

NOTES

While the initial plan was to classify mentions of fluctuations as relevant/other, and then classify the relevant mentions as mental state/other, the distribution of the classes mental state/other was quite imbalanced, with

81% of the mentions already falling under mental state. This is why no further machine learning was applied to this class, and it can be assumed that most relevant mentions of fluctuations will be related to mental state.

Production

- Status – 'open' or 'owned'
- Run schedule – on request
- Version - 1

27. FORMAL THOUGHT DISORDER

Description

Application to extract occurrences where formal thought disorder is present.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include deteriorating into a more thought disordered state with outbursts of aggression; there was always a degree thought disorder. Also include some formal thought disorder.

Negative annotations include thoughts: no FTD, no signs of FTD, NFTD.

Unknown annotations include '?FTD', 'relative shows FTD', 'check if FTD has improved', used in a list, not applying to patient 'typical symptoms include...'.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term '*ftd*, *formal* *thought* *disorder*)

Search Terms (case insensitive)

ftd

formal *thought* *disorder*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 50 (one document per patient). 20 documents were evaluated on top of the initial 30 to confirm that precision was low (<80%)	P=72%			
2		Random sample of 100 – 3 CCS correspondence-attached text, 3	P=56%	Random sample of 100 - 50 events- clinical notes, 50	P=57%, R=36%	formal thought disorder ftd

		discharge notification summaries, 1 mental state formulation, 1 presenting circumstances, 10 ward progress notes, 38 events- clinical notes, 44 correspondence- attached text		correspondence- attached text		
3		Random sample of 100 – 7 ward progress notes, 3 discharge notification summaries, 4 CCS correspondence- attached text, 1 CAMHS event- clinical note, 56 correspondence- attached text, 29 event- clinical note	P=82%	Random sample of 100 - 50 events- clinical notes, 50 correspondence- attached text	P=57% R=61%	formal thought disorder ftd
4	Application excludes instances of 'NFTD'	Random sample of 100 – 9 CCS correspondence- attached text, 3 body text, 1 discharge notification summary, 1 mental state formulation, 50 correspondence- attached text, 36 events- clinical note	P=85%	Random sample of 100 - 50 events- clinical notes, 50 correspondence- attached text	P=83% R=83%	formal thought disorder ftd

NOTES

False positives include negations - did not display, not displaying, not expressed, no evidence of, without showing, uncertainty - unable to elicit, possible…, not possible to assess. Also, no sign of paranoia or formal thought disorder, without showing clear formal thought disorder.

Code for post-processing

name not like 'NFTD'

Production

- Run schedule – monthly
- Version - 1

28. GRANDIOSITY

Description

Application to extract occurrences where grandiosity is apparent.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include ZZZZZ was wearing slippers and was animated elated and grandiose, few grandiose statements regarding having been 'brought up with royalty'. Also include reduction in grandiosity/no longer grandiose.

Negative annotations include no evidence of grandiose of delusions in the content of his speech, no evidence of grandiose ideas.

Unknown annotations include his experience could lead to grandiose ideas.

Interrater reliability

Cohen's k = 89% (50 un-annotated documents - 25 events/25 attachments, search term 'grandio*')

Search Terms (case insensitive)

grandios

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients, random sample of 30 (one document per patient)	P=97%			
2		Random sample of 100 – 2 ward progress notes, 2 presenting circumstances, 1 mental state formation, 49 correspondenc	P=89%	Random sample of 100 - 50 events-clinical notes, 50 corresponde nce-	P=95%, R=91	grandios*

		e- attached text, 46 events- clinical notes		attached text		

NOTES

The majority of false positives occurred due to negations, e.g. ‘no grandiose delusions’, ‘denied...’, ‘nil…’, ‘no evidence of….’. One unknown mention was due to unsure term ‘some possible’. False negatives occurred when the word grandiose was the first word of the sentence e.g. ‘Grandiose, feels...’ and ‘Grandiose beliefs still expressed’. Perhaps this is to do with the capitalisation of G or simply the order of the terms in the sentence.

Production

- Run schedule – monthly
- Version - 1

29. GUILT

Description

Application to identify instances of guilt.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include she then feels guilty/angry towards mum; being angry is easier to deal with than feeling guilty. Also include feelings of guilt with a reasonable cause and mentions stating 'no longer feels guilty'.

Negative annotations include no feeling of guilt, denies feeling hopeless or guilty.

Unknown annotations include 'he might be feeling guilty', 'some guilt' or 'sometimes feeling guilty', or when used in a list, not applying to patient 'typical symptoms include ….'.

Interrater reliability

Cohen's k = 92% (50 un-annotated documents - 25 events/25 attachments, search term 'guil*')

Search Terms (case insensitive)

guil

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients with primary diagnosis code F32* and F33* in a structured field, random sample of 90 (one document per patient).	P=73%			
2	Application searches free text for instances of 'guilt*' (see notes)	All patients with primary diagnosis code F32* and F33* in a structured field, random	93%			guil*

		sample of 90 (one document per patient).				
3	As above	Random sample of 100 – 1 mental health formulation, 16 ward progress notes, 25 correspondenc e- attached text, 58 events-clinical note	P=81%	Random sample of 100 - 50 events- clinical notes, 50 corresponde nce- attached text	P=78%, R=95%	guilt*
4	As above	Random sample of 100 – 3 ward progress notes, 1 mental health care plan, 2 CCS correspondenc e-attached text, 28 correspondenc e-attached text, 2 CAMHS events-clinical notes, 36 events-clinical notes	P=84%	Random sample of 100 - 50 events- clinical notes, 50 corresponde nce- attached text	P=83% R=83%	guilt*

NOTES

Most of the false positives were due to criminal charges e.g. Plead/pleaded guilty, guilty of charges. Others were guilt presented in the same list form sentence 'anxiety, thoughts of suicide, guilt, hope, self-esteem' or negation, specifically 'denies guilt'. The only pattern seen for false negatives was using the word 'feels' or 'feel' guilty.

Code for post-processing

name like 'guilt%'

Production

- Run schedule – monthly
- Version - 1

30. HALLUCINATIONS (ALL)

Description

Application to identify instances of hallucinations.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: her husband was minimising her hallucinations, continues to experience auditory hallucinations, doesn't appear distressed by his hallucinations, he reported auditory and visual hallucinations, this will likely worsen her hallucinations, his hallucinations subsided, neuroleptics were prescribed for her hallucinations, it is unclear if hallucinations occur within the context of a delirium, has delirium been ruled out as a cause for the hallucinations?, he used to experience hallucinations but not anymore, when she relapses she presents with experience of hallucinations, difficult to assess if hallucinations have gone, visual hallucinations ++

Negative annotations include: denied any hallucinations, no evidence of auditory hallucinations, he reports it is a dream rather than hallucinations, hears voices but denies command hallucinations, did not report any further auditory hallucinations, hallucinations have not recurred, no longer appeared to have hallucinations, has not had hallucinations for the last 4 months, the hallucinations stopped, auditory hallucinations -

Unknown annotations include: probably/possibly/maybe/likely/unclear/unable to ascertain/unconfirmed reports of hallucinations/ experiencing hallucinations, pseudo(-) hallucinations, hallucinations present? ?hallucinations, hallucinations?, this is not a psychiatric symptom such as hallucinations, perceptions: abnormalities including hallucinations, derealisation etc., rating scale including delusions, hallucinations, clinical domains e.g. hallucinations, hallucinations is a sign of relapse, it is unusual for hallucinations to present in this way, CBT is effective for hallucinations.

Interrater reliability

Cohen's k = 83% (100 un-annotated documents - 50 events/50 attachments, search term 'hallucinat*')

Search Terms (case- insensitive)

hallucinat*

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		Random sample of 100 – 39	P=75%	Random sample of 100 – 50	P=88% R=90%	hallucinat*

		Correspondence-attached text, 1 mental health care plan, 2 CCS_correspondence - attached text, 3 discharge notification summaries, 7 ward progress notes,1 risk event description, 1 presenting circumstances		events – comments, 50 correspondence – attached text		
2	Application excludes '*no reported halluc*', '*no evidence of halluc*', '*no halluc*', 'denied halluc*', '*possibly halluc*' (see notes)	Random sample of 100 – 7 ward progress notes, 1 mental state formulations, 2 risk events, 41 correspondence-attached texts, 2 menta health care plans, 2 CCS correspondence-attached text, 1 CCS correspondence-body text, 3 discharge notification summaries,41 events-comments	P=90%	Random sample of 100 – 50 events – comments, 50 correspondence – attached text	P=84% R=98%	hallucinat*

NOTES

Round 1

Regarding false positives, most were due to unknown 'possible' and 'possibly' instances being labelled as a positive, as well as negations such as 'no evidence of', 'no reported' and 'no clear indication'. Regarding false negatives, instances were incorrectly labelled as positive when there was a negation shortly before the hallucination mention.

Round 2

Regarding false positives, the few instances that occurred were a few mentions of ‘nil’, ‘did not experience’ and ‘denying’. Regarding false negatives, there were not enough mentions to decipher a pattern.

Code for post-processing

contextstring not like '%no reported halluc%' and contextstring not like '%no evidence of halluc%' and contextstring not like '%no halluc%' and contextstring not like '%denied halluc%' and contextstring not like '%possibly halluc%'

Production

- Run schedule – monthly
- Version - 2

31. HALLUCINATIONS - AUDITORY

Description

Application to identify instances of auditory hallucinations non-specific to diagnosis.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, negative and unknown.

Positive annotations include Seems to be having olfactory hallucination, in relation to her tactile hallucinations.

Negative annotations include denies auditory, visual, gustatory, olfactory and tactile hallucinations at the time of the assessment; denied tactile/olfactory hallucination.

Unknown annotations include possibly olfactory hallucinations, symptoms include….

Interrater reliability

Cohen's k = 96% (50 un-annotated documents - 25 events/25 attachments, search term 'auditory' or 'halluc*')

Search Terms (case insensitive)

auditory hallucinat*

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		Random sample of 100 – 36 attachments, 2 ccs correspondence, 2 mental health care plans, 6 discharge summaries, 47 events and 7 ward progress notes	P=92%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P= 80%, R=84%	auditory halluc*

NOTES

The majority of false positives occurred when ‘denied/denies’ was used to negate the term ‘auditory hallucinations’. The app correctly annotates the phrase ‘no auditory hallucinations’ as a negative mention. However, the phrase ‘no auditory/visual hallucinations’ is annotated as a positive mention.

Production

- Run schedule – monthly
- Version - 1

32. HALLUCINATIONS – OLFACTORY TACTILE GUSTATORY (OTG)

Description

Application to extract occurrences where auditory hallucination is present. Auditory hallucinations may be due to a diagnosis of psychosis/schizophrenia or may be due to other causes, e.g. due to substance abuse.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, negative and unknown.

Positive annotations include seems to be having olfactory hallucinations, in relation to her tactile hallucinations.

Negative annotations include denies auditory, visual, gustatory, olfactory and tactile hallucinations at the time of the assessment, denied tactile/olfactory hallucinations.

Unknown annotations include possibly olfactory hallucinations, common symptoms include….

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'olfact*' or 'gustat*' or 'tactile')

Search Terms (case insensitive)

olfactory [0-10 words in between] *hallucin*

hallucin [0-10 words in between] *olfactory*

gustat [0-10 words in between] *hallucin*

hallucin [01-10 words in between] *gustat*

tactile [0-10 words in between] *hallucin*

hallucin [0-10 words in between] *tactile*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 50	P=86%			
2		Random sample of 100 –	P=86%	Random sample of	P=78%, R=68%	olfactory

		19 correspondence- attached text, 6 mental health care plan, 2 discharge summaries, 19 CCS correspondence- attached text, 1 mental health formulation, 1 ward progress note, 52 events-clinical notes		100 - 50 events- clinical notes, 50 correspondence- attached text		gustat* tactile

NOTES

False positives were negations e.g. no visual/tactile hallucinations, denied any hallucinations, nil olfactory/gustatory hallucinations. 'Denies' seems to be a common false positive pattern. Unknown mentions were vague terms e.g. 'I wonder', 'it is not clear', or questioning whether the symptoms was present.

Production

- Run schedule – monthly
- Version - 1

33. HALLUCINATIONS - VISUAL

Description

Application to extract occurrences where visual hallucination is present. Visual hallucinations may be due to a diagnosis of psychosis/schizophrenia or may be due to other causes, e.g. due to substance abuse.

Definition

Development approach: Machine-learning**.**

Classification of past or present symptom: Both.

Classes produced: Positive, negative and unknown.

Positive annotations include responding to visual hallucination, experiencing visual hallucination, history of visual hallucination, distressed by visual hallucination

Negative annotations include denied any visual hallucination, not responding to visual hallucination, no visual hallucination, no current visual hallucination (with no reference to past).

Unknown annotations include if/may/possible/possibly/might have visual hallucinations, monitor for possible visual hallucination.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'visual' and 'halluc*')

Search Terms (case insensitive)

visual hallucinat*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 – 8 ward progress notes, 1 mental state formulation, 1 mental state comment, 1 CAMHS event, 2 mental health care plans, 1 discharge notification	P=86%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P=77% R=64%	visual and halluc*

		summary, 3 CCS correspondence- attached text, 46 correspondence-attached text, 37 events-clinical note				
2	Application excludes instances of '*no visual*' and '*or visual*' (see notes)	Random sample of 100 - 4 mental state formulations, 10 ward progress notes, 3 mental health care plans, 2 CCS correspondence-attached text, 2 discharge notification summaries, 31 correspondence-attached text, 48 event-clinical note	P=83%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P=91% R=96%	visual hallucination*

NOTES

The main false positives occurred with the term 'possible visual hallucinations' or 'possible previous visual hallucinations'. Others were vague terms such as 'verging on…', 'not currently having…' with no reference to having it previously. A few negations e.g. 'denies' and 'nil' were also falsely labelled positive.

Code for post-processing

contextstring not like '%no visual%' and *contextstring* not like '%or visual%'

Production

- Run schedule – monthly
- Version - 1

34. HELPLESSNESS

Description

Application to identify instances of helplessness.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive and negative.

Positive annotations include ideas of helplessness secondary to her physical symptoms present, ideation compounded by anxiety and a sense of helplessness, hopelessness.

Negative annotations include denies uselessness or helplessness, no thoughts of hopelessness or helplessness. Include also when nothing stated or 'felt helpless when' statements.

Unknown annotations include is there a sense of helplessness, helplessness is a common symptom.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'helpless*')

Search Terms (case insensitive)

helpless

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients with primary diagnosis code F32* or F33* in a structured field, random sample of 30 (one document per patient).	P=90%			
2		Random sample of 100 – 42 correspondence- attached	P=92%	Random sample of 100 - 50 events- clinical	P=93% R=86%	helpless*

		text, 50 events- clinical note, 2 mental health care plans, 2 presenting circumstances , 4 mental health formulations		notes, 50 corresponde nce- attached text		

NOTES

Half of the false positives that did occur in the annotated documents were due to negations of ‘denies’, while the other half were unknowns e.g. Questioning whether this symptom was occurring.

Production

- Run schedule – monthly
- Version - 1

35. HOPELESSNESS

Description

Application to identify instances of hopelessness.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include feeling very low and hopeless, says feels hopeless.

Negative annotations include denies hopelessness, no thoughts of hopelessness or helplessness.

Unknown annotations include is there a sense of hopelessness, hopelessness is a common symptom.

Interrater reliability

Cohen's k = 90% (50 un-annotated documents - 25 events/25 attachments, search term 'hopeless*')

Search Terms (case insensitive)

hopeles

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients with primary diagnosis code F32* or F33* in a structured field, random sample of 30 (one document per patient).	P=87%			
2		Random sample of 100 – 32 attachment text – attachment, 1 attachment text-CCS_correspon	P=88%	Random sample of 100 - 50 events-clinical notes, 50 corresponde nce-	P=90% R=95%	hopeless*

		dence, 61 comments-events, 1 assessment-summary_comments – mental state formulation, 4 mental state comments-mental state formulation, 1 comment – ward notes		attached text		

NOTES:

The majority of false positives was the negation ‘denies’, with some unknowns being questions asking if the symptom is present.

Production

- Run schedule – monthly
- Version - 1

36. HOSTILITY

Description

Application to identify instances of hostility.

Definition

Development approach: Machine-learning.

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include increased hostility and paranoia, she presented as hostile to the nurses.

Negative annotations include not hostile, denied any feelings of hostility.

Unknown annotations include he may become hostile, hostility is something to look out for.

Interrater reliability

Cohen's k = 94% (50 un-annotated documents - 25 events/25 attachments, search term 'hostil*')

Search Terms (case insensitive)

hostil

Performance

	Post-processing rules added to application	**Annotated documents**	**Performance (annotated)**	**Un-annotated documents from keyword search in CRIS**	**Performance (un-annotated)**	**Keywords used for random extraction from CRIS**
1		All patients, random sample of 30 (one document per patient)	P=87%			
2		Random sample of 100 – 1 ward progress note, 1 event-clinical note, 23 discharge notification summaries, 51 CAMHS event-clinical notes, 13 correspondence-attached text, 22 risk event descriptions	P=86%	Random sample of 100 - 50 events- clinical notes, 50 correspondence-attached text	P=89%, R=94%	hostil*

NOTES

The majority of false positives were negations e.g. Never hostile, not hostile, not in a hostile way, with some unknowns being hostility instances not relating to the patient e.g. Relative being hostile towards the patient.

Production

- Run schedule – monthly
- Version - 1

37. INSOMNIA

Description

Application to identify instances of insomnia.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced - Positive, Negative and Unknown.

Positive annotations include any insomnia described including initial insomnia, middle insomnia, any assumed application to the patient - 'the insomnia', complaining of insomnia, taking X for insomnia, contributes to her insomnia, problems with insomnia, this has resulted in insomnia, this will address his insomnia.

Negative annotations include no insomnia, no evidence of insomnia, not insomniac.

Unknown annotations include typical symptoms include insomnia, might have insomnia, ?insomnia, possible insomnia, monitor for insomnia, insomnia has improved.

Interrater reliability

Cohen's k = 94% (50 un-annotated documents - 25 events/25 attachments, search term 'insomn*')

Search Terms (keywords are case insensitive)

insom

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients with primary diagnosis code F32* or F33* in a structured field, random sample of 50 (one document per patient).	P=83%			
2	Application excludes instances of 'winsome' (see notes)	All patients with primary diagnosis code F32* or F33* in a structured	P=94%			

		field, random sample of 50 (one document per patient).				
3	As above	Random sample of 100 – 2 mental state formulations, 4 ward progress notes, 4 mental health care plans, 46 correspondence-attached text, 44 events-clinical notes	P=97%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P=89%, R=94%	insomn*

NOTES

False positives were some negations that weren't picked up and unknown mentions e.g. no longer keen to join the insomnia group.

Code for post-processing

Name not like 'winsome'

Production

- Run schedule – monthly
- Version - 1

38. IRRITABILITY

Description

Application to identify instances of irritability.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced - Positive, Negative and Unknown.

Positive annotations include can be irritable, became irritable, appeared irritable, complained of feeling irritable.

Negative mentions include no evidence of irritability, no longer irritable, no sign of irritability.

Unknown annotations include irritable bowel syndrome, becomes irritable when unwell, can be irritable if …[NB some ambiguity with positive 'can be' mentions, although linked here with the 'if' qualifier], less irritable.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'irritabil*' or 'irritabl*')

Search Terms (case insensitive)

irritabl

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 – 2 mental state formulations, 15 correspondence-attached text, 37 events-clinical notes, 46 ward progress notes	P=99%		P=100% R=83%	irritabil* irritabl*

NOTES

The only false positive found in the annotated document was an irrelevant mention of irritable bowel syndrome. There was no clear pattern found for false negatives, but that was probably due to their low frequency.

Production

- Run schedule – monthly
- Version - 1

39. LOSS OF COHERENCE

Description

Application to identify instances of incoherence or loss of coherence in speech or thinking.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced - Positive, Negative and Unknown.

Positive annotations include patient was incoherent, his speech is characterised by a loss of coherence.

Negative annotations include patient is coherent, coherence in his thinking.

Unknown annotations include coherent discharge plan, could not give me a coherent account, more coherent, mood was coherent with speech and a few instances where coherence/incoherence was part of a heading or question.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'incoheren*')

Search Terms (case insensitive)

coheren*, incoheren*

Performance

	Post-processing rules added to application	Annotated documents	Performance (annotated)	Un-annotated documents from keyword search in CRIS	Performance (un-annotated)	Keywords used for random extraction from CRIS
1		All patients with primary diagnosis code F32* or F33* in a structured field, random sample of 50 (one document per patient).	P=93%			
2		Random sample of 100 – 16 events-comments, 36 events-comments, 54 correspondenc	P=85%	Random sample of 100 - 50 events-clinical notes, 50 corresponde	Not enough positive annotations	coheren*

		e- attachment text, 52 care plan- outcome detail (47 mental health, 5 physical health		nce- attached text		
3		Random sample of 100 – 16 events- comments, 36 events- comments, 54 correspondence- attachment text, 52 care plan- outcome detail (47 mental health, 5 physical health	P=85%	Random sample of 100 - 50 events- clinical notes, 50 correspondence- attached text	Not enough positive annotations	*coheren*
4		Random sample of 158– 16 events- comments, 36 events- comments, 54 correspondence- attachment text, 52 care plan- outcome detail (47 mental health, 5 physical health	P=85%	Random sample of 100 - 50 events- clinical notes, 50 correspondence- attached text	P=98% R=95%	incoheren*

NOTES

False positives mainly occurred with coheren* search term; classifying speech/communication and thinking as coherent rather than not coherent.

False positives sometimes occurred when irrelevant comments were made, such as a relative being incoherent or when describing the need for a coherent treatment plan.

Undetected terms (and negative instances) suggest that the app may randomly interchange between 'coheren*' and 'incoheren*' as positive or negative.

Production

- Run schedule – monthly
- Version - 1

40. LOW ENERGY

Description

Application to identify instances of low energy.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include low energy, decreased energy, not much energy, no energy.

Negative annotations include no indications of low energy, increased energy.

Unclear annotations include typical symptoms include..., might be caused by low energy, monitor for low energy, energy levels have improved, fluoxetine reduces her energy, some energy, energy bars.

Interrater reliability

Cohen's k = 95% (50 un-annotated documents - 25 events/25 attachments, search term 'energ*')

Search Terms (case insensitive)

energy

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients with a primary diagnosis code F32* or F33* in a structured field, random sample of 50 (one document per patient). 20 documents were evaluated on top of the initial 30 to confirm that precision was low (<80%).	P=76%			

2		Random sample of 100 – 1 ward progress note, 1 physical health care plan, 45 correspondence-attached text, 53 events-clinical notes.	P=87%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P=72% R=67%	energ*
3	Terms: low*, lack*, no, poor, reduced, limited, little, none, less, little, decreased 1. Term is an adjective and is within three words of the energy keyword 2. Term is a determiner and is within three words of the energy keyword 3. Term is a verb and is with three words of the energy keyword 4. Term is a noun (optionally immediately followed by the word “of”), if within three words of the energy keyword 5. Energy keyword is immediately followed by either a = or a – followed by the term.	Sample of 100 CAMHS events	P=89%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P= 82% R=85%	energ*

NOTES

There was no pattern with false positives. The majority related to high energy levels described in different ways e.g. increased energy, good energy levels, fair energy levels, no difficulties with her energy, more energetic. Other false positives were irrelevant mentions e.g. EDF energy, eating energy bars, and using energy on specific tasks. There were a few unknown mentions such as stating the term energy without reporting whether this was lacking or not. False negatives included fatigue impacts energy, decreased energy, not much energy, low energy, no energy.

Production

- Run schedule – monthly
- Version - 1

38. MOOD INSTABILITY

Description

This application identifies instances of mood instability.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include she continues to have frequent mood swings, expressed fluctuating mood.

Negative annotations include no mood fluctuation/no rapid cycling/no mood unpredictability, denied diurnal mood variations.

Unknown annotations include mood changes not specifically indicative of fluctuation like ‘she had harmed others in the past when her mood changed’, tried antidepressants in the past but they led to fluctuations in mood, no change in mood, her mood has not changed and she is still depressed.

Interrater reliability

Cohen's k = 91% (50 un-annotated documents - 25 events/25 attachments, search term ‘mood’)

Search Terms (case insensitive)

Change [0-2 words in between] *mood*

Changeable [0-2 words in between] *mood*

Changeable [0-2 words in between] *mood*

Changes [0-2 words in between] *mood*

Extremes [0-2 words in between] *mood*

fluctuate [0-2 words in between] *mood*

Fluctuates [0-2 words in between] *mood*

Fluctuating [0-2 words in between] *mood*

Fluctuation [0-2 words in between] *mood*

Fluctuations [0-2 words in between] *mood*

Instability [0-2 words in between] *mood*

labile [0-2 words in between] mood

lability [0-2 words in between] mood

Liability [0-2 words in between] mood

Liable [0-2 words in between] mood

Rapid cycling [0-2 words in between] mood

swings [0-2 words in between] mood

unpredictable [0-2 words in between] mood

Unsettled [0-2 words in between] mood

Unstable [0-2 words in between] mood

variable [0-2 words in between] mood

variation [0-2 words in between] mood

volatile [0-2 words in between] mood

Mood [0-2 words in between] change

mood [0-2 words in between] Changeable

Mood [0-2 words in between] Changeable

mood [0-2 words in between] changes

Mood [0-2 words in between] Extremes

Mood [0-2 words in between] fluctuate

Mood [0-2 words in between] Fluctuates

Mood [0-2 words in between] Fluctuating

Mood [0-2 words in between] *mood*

Mood [0-2 words in between] Fluctuations

Mood [0-2 words in between] Instability

Mood [0-2 words in between] *labile*

Mood [0-2 words in between] *lability* Mood [0-2 words in between] Liability

Mood [0-2 words in between] Liable

Mood [0-2 words in between] Rapid cycling

Mood [0-2 words in between] *swings*

Mood [0-2 words in between] *unpredictable*

Mood [0-2 words in between] Unsettled

Mood [0-2 words in between] Unstable

Mood [0-2 words in between] *variable*

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients, random sample	P=72%			

		of 50 (one document per patient). 20 documents were evaluated on top of the initial 30 to confirm that precision was low (<80%).				
2		Random sample of 100 – 17 ward progress notes, 2 mental health care plans, 38 correspondence-attached text, 43 events-clinical notes	P=91%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P=100% R=70%	mood

NOTES

False positives found in the annotated documents were due to negations e.g. 'not labile', 'no complaints of' and hypothetical 'if' situations. Unknown mentions were when a justifiable mood change that was context specific with no mention of general mood instability or consistent mood changes. False negatives were when mood was described as 'fluctuating rapidly' and with 'dips' or violent 'shifts' in mood.

Production

- Run schedule – monthly
- Version - 1

39. MUTISM

Description

Application to identify instances of mutism.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include she has periods of 'mutism', he did not respond any further and remained mute.

Unknown annotations include her mother is mute, muted body language.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'mut*')

Search Terms (case insensitive)

mute

mutism

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 30 (one document per patient).	P=93%			
2		Random sample of 100 – 1 mental state formulation, 6 ward progress notes, 39 correspondence-attached text, 54 events-clinical notes	P=95%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P=91% R=75%	mut*

NOTES

Almost every false positive occurred when the staff surname 'Mutemi' was mentioned. One unknown mention was when a relative of the patient was described as mute. False negatives occurred with the simple term 'mute', no other pattern was seen.

Production

- Run schedule – monthly
- Version - 1

40. NEGATIVE SYMPTOMS

Description

Application to identify instances of negative symptoms.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include she was having negative symptoms, diagnosis of schizophrenia with prominent negative symptoms.

Negative annotations include no negative symptom, no evidence of negative symptoms.

Unknown annotations include are negative symptoms present?, negative symptoms can be debilitating.

Interrater reliability

Cohen's k = 85% (50 annotated documents - 25 events/25 attachments)

Search Terms (case insensitive)

negative *symptom*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 30 (one document per patient).	P=87%			
2		Random sample of 100 – 58 attachments, 41 events	P=87%	Random sample of 100 – 50 attachments, 50 events	P=86% P=95%	negative symptom*

NOTES

Precision and recall are high for both annotated and non-annotated documents. Most mentions of negative symptoms relate to present symptoms (92%). False positives were due to the app failing to identify negation e.g. ‘no negative symptoms’ or due to unknown mentions e.g. ‘possible negative symptoms’ being raised as positive mentions. All false negatives were incidences where ‘N’ was capitalised in ‘Negative symptoms’.

Production

- Run schedule – monthly
- Version - 1

41. NIGHTMARES

Description

Application to identify instances of nightmares.

Definition

Development approach: Machine-learning**.**

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include she was having nightmares, unsettled sleep with vivid nightmares.

Negative annotations include no nightmares, no complains of having nightmares.

Unknown annotations include it's been a nightmare to get this arranged, a nightmare scenario would be….

Interrater reliability

Cohen's k = 95% (50 un-annotated documents - events, search term 'nightmare*')

Search Terms (case insensitive)

nightmare*

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		Random sample of 100 – 25 correspondence-attached text, 11 CAMHS event-comments, 2 CCS correspondence-attached text, 3 discharge notification summaries, 1 mental state formulation, 3 presenting circumstances, 2	P=88%	Random sample of 100 – 50 attachments, 50 events	P=64% R=98%	nightmare*

		ward progress notes, 53 events-comments				
2		Random sample of 100 – 1 presenting circumstance, 6 ward progress notes, 2 mental state formulations, 3 CCS correspondence-attached text, 7 CAMHS events, 36 correspondence-attached text, 45 events-clinical notes	P=93%	Random sample of 100 – 50 attachments, 50 events	P=65% R=100%	nightmare*
3	Application excludes instances of '*nightmare"*', '*nightmare''*', '*no nightmare*', *nil nightmare*', '* “nightmare*', '* “nightmares*', '* “nightmare”*’, '*Nightmare”*', ‘nightmare’*', '*Nightmare’*' (see notes)	Random sample of 100 – 2 mental state formulations, 1 presenting circumstances, 6 ward progress notes, 39 correspondence-attached text, 9 CAMHS event-comments, 1 mental health care plan, 2 CCS correspondence-attached text, 2 discharge notification summary, 39 event-comments	P=89%	Random sample of 100 – 50 attachments, 50 events	P=89% R=100%	nightmare*

NOTES

False positives remain whereby the individual is referring to ‘nightmare’ in a metaphorical sense. Other false positives are due to (more complex) negation problems e.g. no episodes of nightmares, she is not having nightmares, nightmares and flashbacks are denied, he does not have nightmares or flashbacks.

Code for post-processing

contextstring not like '%nightmare"%' and *contextstring* not like '%nightmare''%' and *contextstring* not like '%no nightmare%' and *contextstring* not like '%nil nightmare%' and *contextstring* not like '%“nightmare%' and *contextstring* not like '%“nightmare”%' and *contextstring* not like '%“nightmare”%' and *contextstring* not like '%Nightmare”%' and *contextstring* not like '%‘nightmare’%' and *contextstring* not like '%Nightmare’%'

Production

- Run schedule – monthly
- Version - 1

42. OBSESSIVE-COMPULSIVE SYMPTOMS

Description

Application to identify obsessive-compulsive symptoms (OCS) in patients with schizophrenia, schizoaffective disorder or bipolar disorder

Definition

Development approach: rule-based

Classification of past or present symptom: Both.

***Classes produced*:** Obsessive-compulsive disorder (OCD) column contains value True and False, OCS column contains value Positive and Negative

For instances where there are OCS, filter by OCS = Positive

For instances of OCD, filter by OCD = True. There should be no cases where OCS is negative but OCD is true

Positive annotations of OCS include

- Text states that patient has OCD features/symptoms
- Text states that patient has OCS
- Text including hoarding, which is considered part of OCS, regardless of presence or absence of specific examples
- Text states that patient has either obsessive or compulsive or rituals or Yale-Brown Obsessive Compulsive Scale (YBOCS) [see keywords below] and one of the following:
    - Obsessions or compulsions are described as egodystonic
    - Intrusive, cause patient distress or excessive worrying/anxiety
    - Patient feels unable to stop obsessions or compulsions
    - Patient recognises symptoms are irrational or senseless
- Clinician provides specific YBOCS symptoms
- Text reports that patient has been diagnosed with OCD by clinician

Negative annotations of OCS include

- Text makes no mention of OCS
- Text states that patient does not have OCS
- Text states that patient has either compulsions or obsessions, not both, and there is no information about any of the following:
    - Patient distress
    - Obsessive or compulsive symptoms described as egodystonic
    - Inability to stop obsessions or compulsions
    - Description of specific compulsions or specific obsessions
    - Patient insight
- Text states that non-clinician observers (e.g., patient or family/friends) believe patient has obsessions or compulsions without describing YBOCS symptoms.
- Text includes hedge words (i.e., possibly, apparently, seems) that specifically refers to OCS keywords
- Text includes risky, risk-taking or self-harming behaviours
- Text includes romantic or weight-related (food-related) words that modify OCS keywords

Interrater reliability

Cohen's k = 80% (600 annotated documents for interrater reliability)

Search Terms (case insensitive)

Keywords are defined below:

OCS Keywords	YBOCS Keywords	Patient Insight Keywords
Obses* (Includes variations such as 'obsessive' and 'obsessional')	Clean* (Includes variations such as 'cleaned' or 'cleanliness')	Distres* (Includes variations such as distressed or distressing)
Compul* Includes varations such as Compulsive or compulsively, but specifically excluding "compulsory".	Wash* (Includes variations such as washing or washed)	Unwanted
OCD* (Includes variations such as OCD and O.C.D)	Check* (Includes variation such as checking and checked)	Repugnant
Hoard* (Includes varations such as Hoarding and Hoarded)	Repeat* (Includes varations such as repeatedly or repetitive)	Repulsive
Ritual* (Includes variations such as 'ritualistic' and 'ritually')	Count* (Includes variations such as counted or counting)	Egodystonic
	Order* (Includes variations such as ordered or ordering)	Intrusive
	Counting	Intruding
	Rearrange*(Includes variations such as rearranging or rearranged)	

Exclusion keywords:

Form	Negation	Other Experiencer	Self-Description	Hedge
c - obsessive compulsive	None	Mother/Father	Self-Described	Seem(s)
hoarded materials blocking passages	deny* (includes variations such as denied and denying)	Sister/Brother	He/She describe(s/d)	Possible* (Including variations such as possibility/and possibly
obsessions and compulsions. none	Nil	Parent	Described Him/Herself	Apparent(ly)

Obsessive Compulsive Index (including variations such as o.c.i, oci)	no(t) obses* (includes variations such as obsessed, obsessions and obsessional)	Son/Daughter	Say(s) that	Sound(s) like
	than (an) obses* (includes variations such as obsessed, obsessions and obsessional)	Sibling	told me	
	No History	Family		
	No Evidence	Boy/Girlfriend		
		Partner		
		Husband/Wife		
		qqqqq (a pseudonym for a family member or carer)		

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients, random sample of 100 (39 Attachment; 16 CAMHS Event; 1 CCS_Correspondence; 2 Discharge_Notification_Summary; 35 Event; 1 History; 2 Summary_Of_Need; 3 Ward Progress Note; 1 WardRound)	P=72%			

2						

NOTES

Development Performance

Performance of individual components of the OCS algorithm in the validation set (300 documents) and the performance overall for detecting any OCS (including OCD) across all strings with Precision (positive predictive value) and recall (sensitivity) provided.

Symptom	Precision	Recall
Obsessions	0.73	0.5
Compulsions	0.63	0.83
OCD	1	0.85
Hoard	0.73	0.81
Ritual	1	0.33
Any OCS (including OCD)	0.77	0.67

Production

- Run schedule – On request
- Version - 1

43. PARANOIA

Description

Application to identify instances of paranoia. Paranoia may be due to a diagnosis of paranoid schizophrenia or may be due to other causes, e.g. substance abuse.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include vague paranoid ideation, caused him to feel paranoid.

Negative annotations include denied any paranoia, no paranoid feelings.

Unknown annotations include relative is paranoid about me, paranoia can cause distress.

Interrater reliability

Cohen's k = 92% (100 annotated documents - 25 events/69 attachments/1 mental state formulation/3 presenting circumstances/2 progress notes)

Search Terms (case insensitive)

paranoi

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients, random sample of 50 (one document per patient).	P=82%			
2		Random sample of 100 - 69 correspondence-attached text, 2 ward progress notes, 3 presenting circumstances	P=89%	Random sample of 100 – 50 attachments, 50 events	P=86%, R=94%	paranoi*

		, 1 mental state formulation, 15 event-clinical notes				

NOTES

Overall precision for annotated documents was 89% but precision was notably higher in attachment documents (94%) than events (72%). This appears to be due to lack of negation terminology used in attachments (0 negations) compared to events (7 negated sentences). This may be because events are referring to the present symptomatology whilst attachments are summarising broader periods of time. As around 30% of app raises are of 'Paranoid Schizophrenia' diagnoses, this app should perhaps only be used for paranoia relating to schizophrenia, rather than for example, dementia or substance misuse. False positives almost exclusively occurred when the app failed to pick up a negation. All negative mentions were annotated as positive suggesting there is no rule for negation. 5/6 false negatives were in the format 'Diagnosis: Paranoid schizophrenia' so may relate to presence of the colon.

Production

- Run schedule – monthly
- Version - 1

43. PASSIVITY

Description

Application to identify instances of passivity.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include presence of passivity in the present admission, or if the symptom is absent currently but has existed in the past. For example, "patient describes experiencing passivity" or "patient has experienced passivity in the past but not on current admission".

Negative annotations include "denies passivity" or "no passivity".

Unknown annotations include passivity stated as not having been explored, if it is unsure whether symptom is in fact present or if the symptom was not fully delineated. For example: "passivity could not be discussed", "possible passivity requiring further exploration" or "unclear whether this is passivity or another symptom".

Interrater reliability

Cohen's k = 83% (438 unannotated documents – search term 'passivity')

Search Terms (case insensitive)

passivity

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
		Random sample of 100 – 44 attachment text-attachment, 3 body-ccs correspondence, 6 comments-CAMHS event, 42 comments-event, 2 comments-CAMHS event,	P=82%	Random sample of 100 – 50 attachments, 50 events	P=68% R=73%	passivity

		1 current problem – presenting circumstances , 2 mental state comments – mental state formulation				
2	Excludes form titled 'Criminal Justice Mental Health Service Mental Health in Custody (MHiC)'	Random sample of 100 – 50 attachment text-attachment, 4 body-ccs corresponden ce, 42 comments-event, 1 current problem – presenting circumstances , 2 mental state comments – mental state formulation, 1 assessment summary comments – mental state formulation	P=88%	Random sample of 100 – 50 attachments, 50 events	P=89% P=100%	passivity

Production

- Run schedule – on request
- Version - 1

44. PERSECUTORY IDEATION

Description

Application to identify instances of ideas of persecution.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include she was having delusions of persecution, she suffered persecutory delusions, marked persecutory delusions, paranoid persecutory ideations, persecutory ideas present.

Negative annotations include denies persecutory delusions, he denied any worries of persecution, no persecutory delusions, no delusions of persecution, did not report persecutory ideas, no persecutory ideation present etc

Unknown annotations include this might not be a persecutory belief, no longer experiencing persecutory delusions.

Interrater reliability

Cohen's k = 91% (50 un-annotated documents - 25 events/25 attachments, search term 'persecut*')

Search Terms (case insensitive)

[Pp]ersecu*

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		Random sample of 100 - 3 ward progress notes, 8 CCS correspondence-attached text, 71 correspondence-attached text, 18 event-clinical notes	P=85%	Random sample of 100 – 50 attachments, 50 events	P=66% P=94%	persecut*
2	Application excludes	Random sample of 100	P=80%	Random sample of 100	P=80% R=96%	persecut*

	instances of '*No persecutory ideation*', '*No persecutory delusion*', '*No paranoid/persecutory ideation*' (see notes)	- 1 presenting circumstance form, 1 POSProforma form, 9 ward progress note-comments,34 correspondence-attached text, 1 CAMHS event-comments, 1 discharge notification summary, 1 CAMHS event, 52 event-clinical note		– 50 attachments, 50 events		

NOTES

Precision was consistent in both annotated and un-annotated documents. False positives were mainly due to the negation ‘denies’ and ‘denied’ but there were other negations raised e.g. ‘no evidence’, ‘nil’, ‘no clear’, and ‘no…/persecution’. Other false positives were relating to actual persecutions of the patient or patients’ family and unknown mentions e.g. possibly/likely/suggestive of persecutory delusion.

Code for post-processing

contextstring not like '%No persecutory ideation%' and *contextstring* not like '%No persecutory delusion%' and *contextstring* not like '%No paranoid/persecutory ideation%'

Production

- Run schedule – monthly
- Version - 1

45. POOR APPETITE

Description

Application to identify instances of poor appetite (negative annotations).

Definition

This app identifies negative mentions of good appetite.

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations applied to adjectives implying a good or normal appetite: fine; OK; reasonable; alright; preserved; satisfactory. Often described in combination with other symptoms (e.g. sleep and appetite normal; sleep and appetite: both preserved).

Negative annotations applied to adjectives implying a poor/declining appetite: loss of; reduced; decrease in; not so good; diminished; lack of; not great. Also, often in combination with other symptoms (poor sleep and appetite; loss of energy and appetite).

'Unknown' annotations include insufficiently informative adjectives: not changed; varies; increased; improving. Also, hypothetical mentions, as a potential side effect, as an early warning sign, as a description of a diagnosis (rather than patient experience), describing a relative rather than the patient, 'appetite suppressants'.

Good appetite and poor appetite will encapsulate the following descriptive terms:

Good or normal appetite (positive)	Poor or reduced appetite
Alright	Absent
Eats well	Decreasing
Eating well	Deficit
Excellent	Diminished
Fine	Gone down
Fair	Loss of
Good	Losing (also loosing)
Has appetite	Lost
Healthy	Low
Intact	Lacking
Not too bad	Lack of
No problem(s)	Lacks
No concern(s)	Less
Not a concern	Not great
No issue(s)	No

Normal	No interest
OK(ay)	Not as good
Preserved	Not very well
Reasonable	Poor
Regular	Reduced
Stable	Reduction
Satisfactory	Small(er)
Steady	Suppress(ed)
Unremarkable	Suppression
Unimpaired	Worse
Denies problems with	Worsening
Denies issues with	

Interrater reliability

Cohen's k = 91% (50 un-annotated documents - 25 events/25 attachments, search term ‘appetite’)

Search Terms (case insensitive)

appetite* [0-3 words in between] *eating* *well

eating* *well* [0-3 words in between] *appetite

appetite* [0-3 words in between]*alright

alright* [0-3 words in between] *appetite

appetite* [0-3 words in between] *eats* *well

eats* *well* [0-3 words in between] *appetite

appetite* [0-3 words in between]*excellent

excellent* [0-3 words in between] *appetite

appetite * [0-3 words in between] *fine

fine* [0-3 words in between] *appetite

appetite * [0-3 words in between] *fair

fair* [0-3 words in between] *appetite

appetite * [0-3 words in between] *good

good* [0-3 words in between] *appetite

appetite * [0-3 words in between] *healthy

healthy* [0-3 words in between] *appetite

appetite * [0-3 words in between] *intact

intact* [0-3 words in between] *appetite

appetite * [0-3 words in between] *not* *too* *bad

not* *too* *bad* [0-3 words in between] *appetite

appetite * [0-3 words in between] *problem

problem* [0-3 words in between] *appetite

appetite * [0-3 words in between] *noproblem**

no* *problem* [0-3 words in between] *appetite

appetite * [0-3 words in between] *not* *a* *concern

not* *a* *concern* [0-3 words in between] *appetite

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients with primary diagnosis code F32* or F33* in a structured field, random sample of 30 (one document per patient)	P=83%			
	Application excludes instances of ‘good’, ‘normal’, ‘fine’, ‘healthy’, ‘reasonable’, ‘ok’, ‘fair’, ‘alright’ (from the negative annotations – see notes)	All patients with primary diagnosis code F32* or F33* in a structured field, random sample of 30 (one document per patient)	P=97%			
2	As above	Random sample of 100-33 correspondenc	P=89%	Random sample of 100 – 50	P=83% R=71%	appetite

		e- attached text, 1 mental health care plan, 1 discharge notification summary, 4 ward progress notes, 1 mental state formulation, 60 event- clinical note		attachments, 50 events		

NOTES

Code for post-processing

Name not like 'good', 'normal', 'fine', 'healthy', 'reasonable', 'ok', 'fair', 'alright'

Production

- Run schedule – monthly
- Version - 1

46. POOR CONCENTRATION

Description

Application to identify instances of poor concentration.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include my concentration is still poor, she found it difficult to concentrate. Also include he finds it hard to concentrate.

Negative annotations include good attention and concentration, participating well and able to concentrate on activities Also include when concentrate is adequate or reasonable.

Unknown annotations include 'gave her a concentration solution; talk concentrated on her difficulties; urine is concentrated. Include when unclear- e.g. 'he is able to distract himself by concentrating on telly'. Include when also states 'improved concentration/able to concentrate better.'

Interrater reliability

Cohen's k = 95% (100 annotated documents – 45 attachments/3 CAMHS events/1 CCS correspondence/35 mental state formulation/1POSProforma/10 ward progress note)

Search Terms (case insensitive)

concentrat

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients with primary diagnosis code F32* and F33* in a structured field, random sample of 50 (one document per patient). 20 documents were evaluated on top of the initial 30 to confirm that	P=76%			

		precision was low (<80%).				
2		Random sample of 100 - 45 correspondence -attached text, 3 CAMHS events-clinical note, 1 CCS correspondence , 1 POSproforma note, 5 mental state formulation, 45 events-clinical notes	P=74%	Random sample of 100 – 50 attachments, 50 events	P=71% R=64%	concentrat*
3	Application excludes instances of concentrat%*, '*concentration good*'	Random sample of 100 - 7 ward progress note, 1 mental state formulation, 3 CAMHS event-clinical note, 1 mental health care plan, 48 correspondence - attached text, 40 event-clinical note	P=88%	Random sample of 100 – 50 attachments, 50 events	P=84% R=60%	concentrat*

NOTES

False negatives included struggled to concentrate, unable to concentrate, lacked concentration and concentration is impaired.

Code for post-processing

Name not like '%good concentrat%' and n*ame* not like '%concentration good%'

Production

- Run schedule – monthly
- Version - 1

47. POOR EYE CONTACT

Description

Application to identify instances of poor eye contact.

Definition

Development approach: Machine-learning**.**

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive mentions include: 'looked unkempt, quiet voice, poor eye contact', 'eye contact was poor', 'she refused eye contact', 'throughout the conversation she failed to maintain eye contact', 'unable to engage in eye contact', 'eye contact was very limited', 'no eye contact and constantly looking at floor'

Negative mentions include: 'good eye contact', 'he was comfortable with eye contact', 'she made eye contact whilst talking', 'excessive eye contact was made throughout our conversation', 'ZZZZZ made occasional eye contact with me', 'eye contact was inconsistent', 'Mr ZZZZZ made reasonable eye contact', 'low voice, average eye contact'.

Unknown mentions: 'she showed increased eye contact', 'I noticed reduced eye contact today'

Interrater reliability

Cohen's k = 92% (100 annotated documents)

Search Terms (case insensitive)

Available on request

Performance

	Post-processing rules added to application	Annotated documents	Performance (annotated)	Un-annotated documents from keyword search in CRIS	Performance (un-annotated)	Keywords used for random extraction from CRIS
1		All patients, random sample of 100 (one document per patient – 32 attachments, 4 CAMHS events, 1 CCS correspondence, 1 discharge notification summary, 45 events, 2 mental state formulation, 1	P=88%	Random sample of 100 – 50 attachments, 50 events	P=81% R=65%	eye contact

		presenting circumstances, 3 single generic assessments, 1 summary of need, 8 progress ward notes, 2 ward rounds				

NOTES

False positives- good eye contact, reasonable eye contact, appropriate eye contact.
False negatives- poor intermittent eye contact and various singular phrases relating to some eye contact.

Production

- Run schedule – monthly
- Version - 1

48. POOR INSIGHT

Description

Applications to identify instances of poor insight.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotation – An instance is classed as positive if the patient's insight is minimal or absent. For example, records which contain a description of insight relating to the words below would be considered negative:

· Lacking/ Lack of

· Doesn't have

· No/ None

· Poor

· Limited

· Insightless

· Absent

· Impaired

· Little

· Loss/ Lost

Negative annotation – An instance is classed as negative if the patient displays a moderate or high degree of insight into their illness. This includes records containing, for example, the following keywords pertaining to insight:

· Clear

· Had/ Has

· Improving

· Partial

· Some

· Good

· Insightful

· Present

· Aware

· Intact

· Reasonable

Unknown annotation – An instance is classed as unknown if:

· There is a lengthy and unclear description of the patient's insight, without a final, specific verdict.

· Insight was not assessed.

· The word 'insight' is not used in a psychiatry context, rendering it irrelevant.

· The record does not refer to the patient's current level of insight, perhaps mentioning predicted/ previous levels instead.

· It doesn't contain the above keywords, despite the general conclusion that can be drawn from it, as this would decrease the overall accuracy of the app.

· Lack of insight not suggestive of psychotic illness, e.g. 'lack of insight into how his drinking affects his son' or 'lack of insight into how she repeats the same cycles with romantic partners'

Interrater reliability

Cohen's k = 88% (50 un-annotated documents - 25 events/25 attachments, search term 'insight*')

Search Terms (case insensitive)

insight

Performance

	Post-processing rules added to application	**Annotated documents**	**Performance (annotated)**	**Un-annotated documents from keyword search in CRIS**	**Performance (un-annotated)**	**Keywords used for random extraction from CRIS**
1		All patients, random sample of 30 (one document per patient)	P=83%			
2		Random sample of 100 - 52 correspondence - attach text, 1 ccs correspondence, 1 discharge summary, 3 mental health care plan, 42 events and 1 mental health formulation	P=85%	Random sample of 100 – 50 attachments, 50 events	P=87% R=70%	insight*

NOTES

False positives often occurred when the term ‘insight’ was at the start of the sentence e.g. Insight: knows he has... or insight: has some understanding.... Unknown mentions were when insight was discussed or suggested a focus point for intervention without direct mention of the patient lacking in insight. There was no clear pattern for false negatives, the terms ’limited’, ‘poor’, ‘lacking’ and ‘insightless’ were often classed as false negatives. However, there were not enough for a distinguished pattern to be made.

Production

- Run schedule – monthly
- Version - 1

49. POOR MOTIVATION

Description

This application aims to identify instances of poor motivation.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Examples of 'positive' annotations include 'poor motivation', 'unable to motivate' self, 'difficult to motivate' self, 'struggling with motivation'. A sizeable number of statements include motivation in a list of deficiencies (e.g. 'poor sleep, appetite, concentration and motivation'). Includes statements about poor motivation for particular activities (although a statement about a patient lacking the motivation to harm himself was categorised as 'unknown').

Negative annotations include any statements implying some motivation in the patient – e.g. includes specific statements that the patient has good general motivation, but also that they are described as motivated to participate in a group, participate in alcohol rehabilitation. Included positive-indicating trajectories (e.g. 'more motivated', 'improving motivation') but only when they described the patient experience (i.e. not describing interventions aiming to improve motivation).

Unknown annotations included some headings like 'Motivation and Performance', tasks/groups designed for motivation, comments about motivation but not clearly indicating whether this was high or low (e.g. variable motivation), plans to ascertain motivation levels, other use of the word (e.g. 'racially motivated'), 'motivating factors'.

Interrater reliability

Cohen's k = 88% (50 un-annotated documents - 25 events/25 attachments, search term 'motiv*')

Search Terms (case insensitive)

lack [word][word] motivat*

Poor [word][word] motivat*

Struggl [word][word] motivate*

no [word][word] motivat*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 30 (one	P=87%			

		document per patient)				
2		Random sample of 100 - 50 CAMHS event comments, 50 correspondenc e- attach text, 50 care plan outcome detail (49 MH, 1 physical health	P=95%	Random sample of 100 – 50 attachments, 50 events	P=85% P=45%	motiv*
3		Random sample of 100 - 50 CAMHS event comments, 50 correspondenc e- attach text, 50 care plan outcome detail (49 MH, 1 physical health	P=95%	Random sample of 100 – 50 attachments, 50 events	P=95% R=38%	*motiv*

NOTES

False positives often occurred when comments were hypothetical and did not reflect actual motivation level. False positives sometimes occurred when motivation related to relatives of the patient rather than the patient themselves. False positives also occurred occasionally when comment stated 'more motivation'. Despite the rule that poor motivation of self-harm should be 'unknown', there were instances where this was still classified as positive. When including evidence of 'present' symptomatology undetected, precision drops from 95.3% to 89.3%.

Production

- Run schedule – monthly
- Version - 1

50. POVERTY OF SPEECH

Description

Application to identify poverty of speech.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include he continues to display negative symptoms including blunting of affect, poverty of speech, he does have negative symptoms in the form of poverty of speech. Also include 'some poverty of speech' and 'less poverty of speech'.

Negative annotations include no poverty of speech, poverty of speech not observed.

Unknown annotations include poverty of speech is a common symptom of…, ?poverty of speech.

Interrater reliability

Cohen's k = 100% (50 annotated documents - 12 events/32 attachments/5 CCS_correspondence, 1 discharge notification summary)

Search Terms (case insensitive)

*Poverty [0-2 words in between] *speech**

*Impoverish [0-2 words in between] *speech**

**speech* [0-2 words in between] poverty*

**speech* [0-2 words in between] impoverish*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 30 (one document per patient)	P=87%			
		Random sample of 100 patients with a diagnosis of schizophrenia - 56 attachment, 5	P=98%			

		ccs_correspondence, 29 events, 10 ward progress notes				
2		Random sample of 100 – 35 correspondence- attach text, 2 body-ccs_correspondence, 1 brief summary-discharge notification summary, 52 comments-event, 1 mental state comment-mental state formulation, 1 comment, 8 comments-ward progress note	P=88%	Random sample of 100 – 50 attachments, 50 events	P=87% R=85%	impoverished speech poverty of speech

NOTES

Precision is high despite the fact the app has no negative or unknown annotations. This is most likely as in most cases where 'poverty of speech' is documented, it is because the symptom is present.

Production

- Run schedule – monthly
- Version - 1

51. POVERTY OF THOUGHT

Description

Application to identify instances of poverty of thought.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive mentions include poverty of thought was very striking, evidence of poverty of thought etc. Also include 'some poverty of thought' and 'less poverty of thought'.

Negative mentions include no poverty of thought, no evidence of poverty of thought.

Unknown mentions include poverty of thought needs to be assessed, …poverty of thought among other symptoms.

Interrater reliability

Cohen's k = 90% (50 annotated documents)

Search Terms (case insensitive)

poverty *of* *thought*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 30 (one document per patient)	P=83%			
2		Random sample of 100 - 31 attachment text, 2 css correspondence, 9 discharge summaries, 53 events, 5 ward progress notes	P=73%	Random sample of 100 – 50 attachments, 50 events	P=91% R=86%	poverty of thought

3	Application excludes instances of '*no poverty of thought*' (see notes)	Random sample of 100 - 38 attachment text, 4 css correspondence, 2 discharge summaries, 1 mental health care plan, 39 events, 13 ward progress	P=96%	Random sample of 100 - 50 events-comments, 50	P=95%, R=93%	poverty of thought
4	As above	Random sample of 100 patients with schizophrenia- 43 attachment text, 4 css correspondence, 3 discharge summaries, 37 events, 12 ward progress notes, 1 mental state formulation ward progress	P=98%			

NOTES

False positives occurred only with unknown annotations e.g. uncertain terms of possible, possibly, maybe and perhaps poverty of thought.

Code for post-processing

contextstring not like '%no poverty of thought%'

Production

- Run schedule – monthly
- Version – 1

52. PSYCHOMOTOR ACTIVITY (CATEGORISATION)

Description

Application to identify instances of psychomotor activity and determine the level of activity

Definition

Development approach: rule-based

Classification of past or present symptom: Both.

Classes produced: Positive/correct and Negative/incorrect/irrelevant

Positive/correct mentions identifies the level of psychomotor activity (as denoted by the keyword) in the context (as denoted by the contextstring). In addition, psychomotor_activity column correctly states whether the reference to abnormal levels of psychomotor activity in the contextstring.

For example: Keyword: 'psychomotor agitation'; Contextstring: 'patient showed psychomotor agitation'; Negativity: 'No'; psychomotor_activity: 'psychomotor agitation'

Negative/incorrect/irrelevant mentions do not successfully identify the level of activity (as denoted by the keyword) in the context (as denoted by the contextstring). Or an instance of psychomotor activity is noted as negated.

For example: Keyword: 'psychomotor activity'; Contextstring: 'normal psychomotor activity'; Negativity: 'yes'; psychomotor_activity': 'psychomotor activity'

Keyword: 'psychomotor activity'; Contextstring: 'change in psychomotor activity'; Negativity: 'yes'; psychomotor_activity: 'psychomotor activity'

Interrater reliability

N/A

Search Terms (case insensitive)

psychomotor

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients Random sample of 100 – 32 Attachment-Text, 2	P=91%	Random sample of 100 – 50 attachments , 50 events	P=92% R=92%	psychomotor

		Brief_Summary, 58 Comments, 1 Current_Problem, 1 Mental_State_Comments, 2 Mental_State_Examination, 3 MentalState, 1 Summary_Of_Need_Physical_Health				

NOTES

N/A

Production

- Run schedule – monthly
- Version – 1

53. SMELL

Description

Application to identify symptoms of loss of smell.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include she has not recovered her sense of smell since she contracted COVID-19 in May 2021; Complains of loss of smell and loss of tastes

Negative annotations include denies any symptoms of loss of smell; Her mother could not smell the food she made

Unknown annotations include no one else could smell it either; she was unsure whether her smell had been affected

Interrater reliability

Not applicable

Search Terms (case insensitive)

Loss of smell

Lack of smell

Performance

	Post-processing rules added to application	Annotated documents	Performance (annotated)	Un-annotated documents from keyword search in CRIS	Performance (un-annotated)	Keywords used for random extraction from CRIS
1		All patients, random sample of 100 – 2 AddictionsEvent, 42 attachment, 4 CAMHS Event, 1 CCS_correspondence, 34 Event, 6 Single-generic_Assessment, 3 Summary_Of_Need, 8 Ward Progress Note	P=92%	Random sample of 100 – 50 attachments, 50 events	P=83 R=100%	Lack of smell Loss of smell

NOTES

Not applicable

Production

- Run schedule – on request

54. SOCIAL WITHDRAWAL

Description

Application to identify instances of social withdrawal.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include she is withdrawn socially from friends and family, Mr ZZZZZ became very isolated and socially withdrawn, some social withdrawal

Negative annotations include not being socially withdrawn, no evidence of being socially withdrawn.

Unknown annotations include social withdrawal is common in depression, need to ask about social withdrawal.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'withdraw*')

Search Terms (case insensitive)

Social [0-3 words in between] withdraw

Withdraw [0-3 words in between] social

Performance

	Post-processing rules added to application	**Annotated documents**	**Performance (annotated)**	**Un-annotated documents from keyword search in CRIS**	**Performance (un-annotated)**	**Keywords used for random extraction from CRIS**
1		All patients, random sample of 30 (one document per patient)	P=90%			
2		Random sample of 100 – 61 correspondence -attached text, 1 CAMHS event, 1 mental health care plan, 2 CCS correspondence , 1 discharge notification summary, 2 ward progress	P=98%	Random sample of 100 – 50 attachments, 50 events	P=60% R=86%	withdraw*

		notes, 1 mental state formulation, 31 events-comments				

NOTES

Differences between positive only and random documents likely due to low number of positive raises found in random documents (6 true positives, 4 false negatives).

Production

- Run schedule – monthly
- Version - 1

55. STUPOR

Description

Application to identify instances of stupor. This includes depressive stupor, psychotic stupor, catatonic stupor, dissociative stupor and manic stupor.

Definition

Development approach: Machine-learning**.**

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include 'ZZZZ presented in a psychotic stupor', 'man with stuporous catatonia', 'he is in a depressive stupor', 'his presentation being a schizoaffective stupor', 'periods of being less responsive/stuporous', 'standing in a stupor'.

Negative annotations include statements which suggest psychiatric stupor is not indicated e.g. not in the state of stupor, presentation not suggestive of depressive stupor, reported not feeling stuporous.

Unknown annotations include annotations include unclear or hypothetical statements such as uncertain statements regarding the patients state such as: ?manic stupor, possible psychotic stupor however need to exclude medical cause and stupors induced by substance abuse such as: drink himself to stupor, drinking heavily and ending up stuporific, drinking to a stupor, drunken stupors.

Interrater reliability

Cohen's k = 96% (50 un-annotated documents - 25 events/25 attachments, search term 'aggress*')

Search Terms (case insensitive)

Stupor*

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		Random sample of 100 - 14 ward progress notes, 2 mental state formulations, 2 presenting circumstances, 2 discharge notification summaries, 1 CAMHS event-	P=88%	Random sample of 100 – 50 attachments , 50 events	P=88% R=87%	stupor*

		clinical note, 2 mental health care plans, 25 correspondence -attachment, 5 CCS correspondence - attached text, 46 correspondence -attached text				

NOTES

Most of the false positives were instances of a stupor due to alcohol. Some were stupor mentions due to medication and other times simple negation e.g. Not a depressive stupor. Unknown mentions were vague terms e.g. related to stupor, may be..., almost stuporous, borderline stupor. There was no direct pattern regarding the false negatives due to the low frequency of them. Most examples of the false negatives are: 'developing depressive stupor’, ‘woke in a stupor’, ‘with … and stupor’, ‘reaction (stupor)’, ‘becoming stuporous’, ‘short periods of stupor’.

Production

- Run schedule – on request
- Version - 1

56. SUICIDAL IDEATION

Description

Application to identify instances of suicidal ideation - thinking about, considering, or planning suicide.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Examples of positive annotations:

1) Her main concerns were his low mood QQQQQ suicidal ideation

2) He has recently sent a letter to mom describing suicidal ideation.

3) QQQQQ then advised of suicidal ideation.

Examples of negative annotations:

1) There was no immediate risk in relation to self-harm or current suicidal ideation.

2) There has been no self-harm and no suicidal ideation disclosed to QQQQQ.

3) She denies having self-harming or suicidal ideation although sometimes would rather sleep and not get up in the morning.

Examples of unknown annotations:

1) Suicidal ideation is a common symptom in depression.

2) It wasn't certain if she was experiencing suicidal ideation.

Interrater reliability

Cohen's k = 92% (50 un-annotated documents - 25 events/25 attachments, search term 'ideation')

Search terms (case insensitive)

suicide ideat*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 30 (one document per patient)	P=97%			

2		Random sample of 100 – CAMHS events	P=87%	Random sample of 100 – 50 attachments, 50 events	P=81% R=87%	ideation

NOTES

False positives mainly occurred with negations e.g. ‘did not/has not expressed ideation’, ‘denies ideation’, ‘… was not an ideation’. Other negatives were irrelevant comments e.g. persecutory, psychotic or paranoid ideation. Unknowns were often uncertain statements where ideation was questioned or vague comments where it could not be deciphered. 83.9% of positives were present suicide ideation, 16.1% were past suicide ideation (stating no ideation currently or no comment on current ideation, only past).

Production

- Run schedule – on request
- Version - 1

57. TANGENTIALITY

Description

Application to identify instances of tangentiality.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include he was very tangential lacked goal directed thinking, there was evidence of tangential speech.

Negative annotations include no evidence of formal thought disorder or tangentiality of thoughts. However, there was no overt tangentiality or loosening of associations.

Unknown annotations include there can be tangentiality, FTD is characterised by tangentiality, go off on a tangent.

Interrater reliability

Cohen's k = 81% (50 un-annotated documents - 25 events/25 attachments, search term 'tangent*')

Search Terms

tangent

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients, random sample of 30 (one document per patient)	P=97%			
2		Random sample of 100 - 5 ward progress notes, 2 mental state forms, 51 events- clinical notes, 1 CCS correspondence -attached text, 41	P=90%	Random sample of 100 – 50 attachments, 50 events	P=99% R=90%	tangent*

		correspondence -attached text				

NOTES

False positives usually occurred with the negation ‘no evidence of’, as well as a few ‘no tangential’ mentions. One unknown mention was when the patient was talking about going off on a tangent. False negatives occurred with the term going off on tangents and tangential thoughts/in his thoughts.

Production

- Run schedule – monthly
- Version - 1

56. TASTE

Description

Application to identify symptoms of loss of taste within community populations.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include a reference to symptoms/experiences of anosmia. For example, 'the patient reported loss of enjoyment of food due to loss of taste' or 'COVID symptoms present such as loss of taste'.

Negative annotations include no reference to symptoms/experiences of loss of taste. For example, 'the patient denied loss of taste', or 'patients' mother reported loss of taste due to COVID'.

Unknown annotations include form when there is reference of loss of taste in terms of an automated letter or email between colleagues or when it is not clear if the patient has symptoms/experiences of loss of taste. For example, 'the patient is not sure if he has lost his taste', or 'don't come to the practice if you have any COVID symptoms such as loss of taste etc'.

Interrater reliability

NA

Search Terms

Loss of taste*, lack of taste*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 100 – 1 AddictionsEvent, 29 attachment, 6 CAMHS Event, 56 Event, 1 Mental_state_formulation, 2 Single-generic_Assessment, 2 Summary_Of_Need, 1 Ward	P=88%	Random sample of 100 – 50 attachments, 50 events	P=95% R=90%	Lack of taste Loss of taste

		Progress Note, 2 WardRound				

NOTES

NA

Production

- Run schedule – monthly
- Version - 1

57. TEARFULNESS

Description

Application to identify instances of tearfulness.

Definition

Development approach: Machine-learning.

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include appeared tearful; was tearful (including was XX and tearful; was tearful and YY); became tearful; moments of tearfulness; a bit tearful.

Negative annotations include not tearful; no tearfulness; denies feeling tearful; no tearful episodes.

'Unknown' annotations were mostly ambiguous statements (e.g. less tearful; couldn't remember being tearful) and statements applying to another person (e.g. mother was tearful) or a person who was not clearly enough the patient.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'tearful*')

Search Terms (case insensitive)

tearful

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 30 (one document per patient)	P=100%			
2		Random sample of 100 - 3 mental state formulations, 1 risk event, 22 correspondence -attached text, 33 ward progress notes, 41 events-clinical notes	P=94%	Random sample of 100 – 50 attachments, 50 events	P=100% R=94%	tearful*

NOTES

False positives usually occurred due to irrelevant mentions of relatives being tearful. Only three other false positives occurred, due to the negation 'not tearful'. There were also very few false negatives, too few to see a pattern. False negatives were often being tearful, tearful at times, can be tearful, became tearful.

Production

- Run schedule – monthly
- Version - 1

58. THOUGHT BLOCK

Description

Application to identify instances of thought block.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include showed some thought block, thought block and paucity of thought.

Negative annotations include denies problems with thought block, no thought block elicited.

Unknown annotations thought block can be difficult to assess, …among thought block and other symptoms.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'thought block*')

Search Terms

thought *block*

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients, random sample of 30 (one document per patient)	P=93%			
2		Random sample of 100 - 7 ward progress notes, 3 mental state formulations, 2 discharge summaries, 33 correspondence -attached text, 55 events-clinical notes	P=92%	Random sample of 100 – 50 attachments, 50 events	P=91% R=75%	thought block*

NOTES

The majority of false positives were of the negation denied/denies, others being: no evidence of, no sign of, did not appear/appear to be thought blocked. Unknown mentions were when the symptom was questioned, or it was suggested as a possible symptom. Regarding false negatives, there was no pattern observed. Mentioned included: ...is thought blocked, presents as thought blocked, thought blocking at times, past experiences of thought block, is thought blocked.

Production

- Run schedule – monthly
- Version - 1

59. THOUGHT BROADCAST

Description

Application to identify instances of thought broadcasting.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include presence of thought broadcast in the present admission, or if the symptom is absent currently but has existed in the past. For example, "patient describes experiencing thought broadcasting" or "patient has experienced thought broadcasting in the past but not on current admission".

Negative annotations include "denies thought broadcasting" or "no thought broadcasting".

Unknown annotations include thought broadcast stated as not having been explored, if it is unsure whether symptom is in fact present or if the symptom was not fully delineated. For example: " thought broadcasting could not be discussed", "possible thought broadcasting requiring further exploration" or "unclear whether this is thought broadcasting or another symptom".

Interrater reliability

Cohen's k = 94% (95 unannotated documents – search term 'thought broadcast*')

Search Terms (case insensitive)

Though* [0-2 words] broadcast*

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
		Random sample of 100 –	P=84%	Random sample of 100 – 50 attachments, 50 events	P=86% R=92%	thought broadcast*

Production

- Run schedule – on request
- Version - 1

60. THOUGHT INSERTION

Description

Application to identify instances of thought insertion.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include presence of thought insertion in the present admission, or if the symptom is absent currently but has existed in the past. For example, "patient describes experiencing thought insertion" or "patient has experienced thought insertion in the past but not on current admission".

Negative annotations include "denies thought insertion" or "no thought insertion".

Unknown annotations include thought insertion stated as not having been explored, if it is unsure whether symptom is in fact present or if the symptom was not fully delineated. For example: "t thought insertion could not be discussed", "possible thought insertion requiring further exploration" or "unclear whether this is thought insertion or another symptom".

Interrater reliability

Cohen's k = 97% (96 unannotated documents – search term 'thought insert*')

Search Terms (case insensitive)

Though* [0-2 words] insert*

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
		Random sample of 100 –	P=84%	Random sample of 100 – 50 attachments, 50 events	P=81% R=96%	thought insert*

Production

- Run schedule – on request
- Version - 1

61. THOUGHT WITHDRAWAL

Description

Application to identify instances of thought withdrawal.

Definition

Classification of past or present: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include presence of thought withdrawal in the present admission, or if the symptom is absent currently but has existed in the past. For example, "patient describes experiencing thought withdrawal" or "patient has experienced thought withdrawal in the past but not on current admission".

Negative annotations include "denies thought withdrawal" or "no thought withdrawal".

Unknown annotations include thought withdrawal stated as not having been explored, if it is unsure whether symptom is in fact present or if the symptom was not fully delineated. For example: "thought withdrawal could not be discussed", "possible thought withdrawal requiring further exploration" or "unclear whether this is thought withdrawal or another symptom".

Interrater reliability

Cohen's k = 95% (76 unannotated documents – search term 'thought withdraw*')

Search Terms (case insensitive)

Though* [0-2 words] withdraw*

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
		Random sample of 100 –	P=84%	Random sample of 100 – 50 attachments, 50 events	P=90% R=88%	thought withdraw*

Production

- Run schedule – on request
- Version - 1

62. WAXY FLEXIBILITY

Description

Application to identify instances of waxy flexibility. Waxy flexibility is a psychomotor symptom of catatonia as associated with schizophrenia, bipolar disorder, or other mental disorders which leads to a decreased response to stimuli and a tendency to remain in an immobile posture.

Definition

Development approach: Machine-learning**.**

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include she presents as catatonic with waxy flexibility, exhibiting waxy flexibility.

Negative annotations include no waxy flexibility, no evidence of waxy flexibility.

Unknown annotations include his right pre-tibial region was swollen and waxy and slightly pink, waxy flexibility is a very uncommon symptom.

Interrater reliability

Cohen's k = 96% (50 un-annotated documents - 25 events/25 attachments, search term 'waxy')

Search Terms

waxy

Performance

	Post-processing rules added to application	**Annotated documents**	**Performance (annotated)**	**Un-annotated documents from keyword search in CRIS**	**Performance (un-annotated)**	**Keywords used for random extraction from CRIS**
1		All patients, random sample of 30 (one document per patient)	P=90%			
2		Random sample of 100 - 14 ward progress notes, 3 CAMHS events, 2 CCS correspondence, 37 correspondence-attached text,	P=81%	Random sample of 100 – 50 attachments, 50 events	P=80% R=86%	waxy

		44 events- clinical note				

NOTES

False positives were often due to irrelevant mentions of waxy e.g. Complexion or ear wax. Other false positives were due to negations e.g. waxy flexibility- 0, no evidence of, no ... or waxy flexibility. Unknown mentions were due to uncertain comment e.g. Maybe/possibility waxy flexibility. There was no apparent pattern with the false negatives, apart from most of them just including the word waxy but implying waxy flexibility. Some of the instances were waxy in her facial movements and posture, and waxy non-responsive presentation.

Production

- Run schedule – monthly
- Version - 1

63. WEIGHT LOSS

Description

Application to identify instances of weight loss.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include significant weight loss, pleased with his weight loss.

Negative annotations include no weight loss; denies weight loss.

Unknown annotations include to maintain adequate dietary intake and avoid weight loss, the latter reduced in line with weight loss.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'weight* loss', 'loss* weight')

Search Terms

Loss [0-2 words in between] *weight*

Lost [0-2 words in between] *weight*

Weight* [0-2 words in between] loss

Weight* [0-2 words in between] lost

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients with primary diagnosis code F32* or F33 in a structured field, random sample of 30 (one document per patient)	P=97%			

2		Random sample of 100 - 100 CAMHS events	P=79%	Random sample of 100 – 50 attachments , 50 events	P=79%, R=92%	weight* loss loss* weight
3	Application excludes instances of '*no signs of weight loss*' '*denied weight loss%* '*no weight loss*'	Random sample of 100 - 6 comments, 4 CCS correspondenc e- attached text, 37 correspondenc e- attached text, 47 event- clinical notes, 3 mental health care plan, 1 risk event, 2 mental state formulation	P=80%	Random sample of 100 – 50 attachments , 50 events	P=90% R=88%	weight* loss loss* weight

NOTES

Many of the false positives were unknown mentions, using uncertain terms such as ‘apparently’ and ‘might’ being used. These also included plans to lose weight or being on a diet with no mention of the effects being current weight loss. Negation examples were: hasn’t lost weight, no weight loss, did not believe she had lost weight or mention of weight gain.

Code for post-processing

contextstring not like '%no signs of weight loss%' and *contextstring* not like '%denied weight loss%' and *contextstring* not like '%no weight loss%'

Production

1. Run schedule – monthly
2. Version – 1

64. WORTHLESSNESS

Description

Application to identify instances of worthlessness.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include feeling worthless, feels hopeless and worthless.

Negative annotations include no worthlessness, denies feelings of worthlessness.

Unknown annotations include his father had told him that he was worthless, would call them worthless.

Interrater reliability

Cohen's k = 82% (50 un-annotated documents - 25 events/25 attachments, search term 'worthless*')

Search Terms (case insensitive)

worthless

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients with primary diagnosis code F32* or F33* in a structured field, random sample of 30 (one document per patient).	P=90%			
2		Random sample of 100 - 2 mental state formulations, 6 ward progress notes, 3 discharge summaries, 1 mental health care plan, 37 correspondence	P=91%	Random sample of 100 – 50 attachments, 50 events	P=88% P=86%	worthless*

		-attached text, 51 events-clinical notes				

NOTES

The majority of false positives occurred due to the negation ‘denies’ and ‘denied’ worthlessness. There were very few unknown mentions.

Production

- Run schedule – monthly
- Version - 1

PHYSICAL HEALTH CONDITIONS

1. ASTHMA

Description

Application to identify patients with diagnosis of asthma.

Definition

Development approach: sem-EHR***.***

Classification of past or present symptom: Both.

Classes produced: Positive.

Positive mentions include:

'past medical history: eczema, asthma', 'diagnosed with asthma during childhood', 'uses inhaler to manage asthma symptoms', 'suffered from an asthma attack', 'ZZZZZ suffers from severe asthma', 'Mrs ZZZZZ has mild asthma'.

Interrater reliability

Cohen's k = 98% (50 patients from patient level testing, 50 documents from annotation level testing, search term 'asthma')

Search Terms

Ontology available on request

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1	The application excludes the following phrases: 'possibl* asthma' 'formcheckbox asthma' 'possibility [0-5 words] asthma' 'copd, asthma, bronchitis, etc'			Random sample of 100	Patient-level P=95% R=84%	asthma

	'diabetes, asthma, injuries, illnesses'					

NOTES

There was no clear pattern of failure with remaining false positives. These included unidentified forms, information sheets, confusion between anxiety and asthma symptoms and the use of 'asthma' as an example of a physical health condition.

There was no clear pattern of failure for false negatives. Examples included:

'Diagnosed with asthma as a child'

'He suffers with asthma'

'how she could manage her asthma better'

'Past medical history – Asthma'

'uses inhalers for asthma'

Production

- Run schedule – monthly
- Version - 1

2. BRONCHITIS

Description

Application to identify patients with diagnosis of bronchitis.

Definition

Development approach: sem-EHR*.*

Classification of past or present symptom: Both.

Classes produced: Positive.

Positive mentions include:

'Recently had COPD (chronic obstructive pulmonary disease', 'ZZZZ had chronic bronchitis, 'Past diagnosis: chronic obstructive airway disease', 'physical health history: asthma, bronchitis', 'centrilobular emphysema'.

Interrater reliability

Search Terms

Ontology available on request

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1	The application excludes the following phrases: 'possibility [0-5 words] chronic obstructive pulmonary' 'possibility [0-5 words] bronchitis' 'risk of [0-8 words] stroke' 'possibl* bronchitis' 'possibl* chronic obstructive pulmonary' 'possibl* copd'	Random sample of 200	P=85%			

	‘suspected chronic obstructive pulmonary’ ‘formcheckbox copd’ ‘suspected copd’’ ‘suspected bronchitis’ ‘formcheckbox chronic obstructive pulmonary’ ‘*exacerbation of chronic obstructive pulmonary disease%state frequency*' ‘copd, asthma, bronchitis, etc’ And FORMS 1, 5, 8 and 10 (details available on request)					
2	As above plus see post-processing rules in Notes			Random sample of 100 for precision and 40 for recall	Patient-level P=85% R=48%	asthma
3	As above			Random sample of 50	Patient-level P=94%	asthma

NOTES

No pattern seen in false negatives. Remaining false positives were due to ‘possibly’ bronchitis mentions.

Post-processing rules

Where phrase_exclude = 0

Where (form_exclude = 0 or (form_exclude = 1 and form_exclude_type like 'form_9%') or (form_exclude = 1 and form_exclude_type like 'form_5%'))

We found common forms and phrases within patient records that reduce precision and were thus excluded from the application. The list of forms is available on request.

Production

- Run schedule – monthly
- Version - 1

3. COUGH

Description

Application to identify instances of coughing.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive examples:

- She has been experiencing a cough for the last week and is going to call her GP.
- ZZZ called ahead of today's session reporting a cough so we agreed to move the session to over the phone due to current COVID guidance.
- He has been to the GP due to coughing up sputum.

Negative examples:

- She denied any coughing or shortness of breath.
- He stated he was unwell with a cold last week, no fever or cough reported.
- Fever, cough, shortness of breath: Nil

Unknown examples:

- She is feeling very distressed because people were coughing near her on the bus
- Her son is currently off school with a bad cough.

Interrater reliability

Cohens k = 79% (150 un-annotated documents, search terms 'cough*')

Search Terms

cough*

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		Random sample of 107	P=78%	Random sample of 99	P=83% R=80%	cough*

2	The application excludes the following phrases: 'no cough' 'nil cough'	Random sample of 100	P=82%			

Production

- Run schedule – monthly
- Version - 1

4. CROHN'S DISEASE

Description

Application to identify patients with diagnosis of Crohn's disease.

Definition

Development approach: sem-EHR***.***

Classification of past or present symptom: Both.

Classes produced: Positive.

Positive mentions include:

'recently been diagnosed with crohn's disease', 'ZZZZ has crohn's disease', 'she has a history of crohn's disease', 'has been hospitalised due to severe crohn's disease', 'physical health history: asthma, diabetes, hypertension, crohn's disease'

Interrater reliability

Cohen's k = 98% (50 patients from patient level testing, 50 documents from annotation level testing, search term 'crohn*')

Search Terms

Ontology available on request

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1	The application excludes the following phrases: 'possibl* crohn*' 'suspected crohn*' 'formcheckbox crohn*' 'risk of [0-8 words] crohn*'			Random sample of 50	Patient level P=94% R=78%	'crohn*'

	'possibility [0-5 words] crohn*' '?crohn*' And FORMS 1 and 5 (details available on request) See post-processing rules in Notes					

NOTES

Remaining false positives occurred in the following instances:

1. Instances where the patient is not the subject:

'son crohns disease'

'her ex partner has crohns disease'

2. Random instances of 'Crohn's' mentioned where the patient does not have a Crohn's disease diagnosis:

'Crohn's diet'

'says she has Crohn's vagina'

There is no clear pattern of failure for remaining false negatives:

'she was diagnosed with Crohn's disease'

'monitor bowels in light of Crohn's diagnosis'

'an elective colonoscopy for Crohn's disease'

'Had surgery in 2011 due to Crohn's disease'

'he said that his Crohn's disease has been acting up'

'history: Psoriasis constipation Crohns social anxiety and depression'

Post-processing rules

Through testing, we have found that optimum precision and recall are indicated when documents containing >2 mentions of the illness are not excluded. We do not exclude documents containing the above phrases if they contain >2 mentions of the illness.

We found common forms and phrases within patient records that reduce precision and were thus excluded from the application. The list of forms is available on request.

Production

- Run schedule – monthly
- Version - 1

5. FALLS

Description

Application to identify instances of falls or falling.

Definition

Development approach: Rules-based**.**

Classification of past or present symptom: Both.

Classes produced

- Fall_single_episode, any reference to a single fall (regardless of when it happened) e.g., 'he fell last night', 'he had one fall 10 years ago').
- Fall_recurrent: any reference to more than one fall (regardless of when they happened), e.g. 'he reported recurrent falls', 'she had a couple of falls.
- Fall_risk: any reference to an actual risk of falling, e.g., "she is at a high risk of falls on account of this medication she is taking"
- Fall_other: to capture any other relevant fall mention that does not belong to one of the previous categories
- Not_relevant: to capture irrelevant mentions or false positives, e.g. 'in the fall', 'falling in love'

Note 1: positive annotations must refer to the patient and not someone else.

His mother had one fall > NOT_RELEVANT

Note 2: hypothetical statements should **not** be counted

If she took this medication, she might be at risk of falling > NOT_RELEVANT

Note 3: classes should be chosen on an annotation level: "She had a fall 10 months ago and then had another fall yesterday" should end up as two single-episode annotations, but "she had a couple of falls: 10 months ago and yesterday" would end up as a recurrent annotation.

Note 4: accidental falls are to be considered relevant

He fell from the bed > FALL_SINGLE_EPISODE, positive

Note 5: mentions where a fall is "suggested" but not explicitly written (e.g. 'Fall pendant', 'Falls clinic', 'Falls referral', 'Falls prevention advice') should be considered as NOT_RELEVANT.

Interrater reliability

Search Terms (case insensitive)

fall*

fell

Performance

	Post-processing rules added to application	Annotated documents	Precision and recall (annotated)	Un-annotated documents extracted	Precision and recall (un-annotated)	Keywords used for

		identified by the application		from keyword search in CRIS		extraction from CRIS
1		Random sample of 100 – 1 addictions event, 356attachments, 1 CAMHS event, 1 care plan mental health, 1 care plan physical health, 2 discharge notifications summaries, 39 events, 1 histories, 1 mental state formulation, 1 presenting circumstances, 2 risk events, 1 risk assessment tool – CRISRiskPlan, 5 risk assessment tools, RiskFactors, 1 summary of need, 5 ward progress notes, 1 ward round	General P=85% Specific category P=69%	Random sample of 100 – 50 attachments, 50 events	P=77% R=58%	fall* fell

NOTES

False positives mainly occurred in negations.

Production

- Run schedule – on request
- Version – 1

6. FEVER

Description

Application to identify patients with any symptom of fever developed within the last month.

Definition

Development approach: machine learning.

Classification of past or present symptoms: past.

Classes produced: positive, negative, and unknown.

Positive examples include:

- She informed me on the phone she has had a fever all week.
- ZZZ has been taking paracetamol for a fever
- Attended A&E reporting fever
- She felt feverish

Negative examples include:

- I asked if she had any symptoms, such as fever, which she denied.
- Temperature was checked for signs of fever, none observed.
- Cough, fever, shortness of breath: Nil

Unknown mentions include:

- Her son had a fever last night and she can’t make it to today’s session.
- She reported worrying over what to do if the baby developed a fever.
- I have informed her that if symptoms worsen, or she develops a fever, to attend A&E.

Search terms: fever. Excluded from the database:

- Yellow fever
- Malaria fever
- Hay fever
- Typhoid fever
- Dengue fever
- Rheumatic fever
- glandular fever
- cabin fever

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1			P = 91%		P = 85% R = 86%	fever*

NOTES

Production

a. Status – ‘open’ or ‘owned’

b. Run schedule: on request

c. Version: 1

7. HYPERTENSION

Description

Application to identify patients with diagnosis of hypertension or high blood pressure.

Definition

Development approach: sem-EHR***.***

Classification of past or present symptom: Both.

Classes produced: Positive.

Positive mentions include:

'Recently been diagnosed with hypertension', 'ZZZZ has high blood pressure', 'she has a history of hypertension', 'physical health history: asthma, diabetes, high blood pressure'.

Interrater reliability

Cohen's k = 91% (50 patients from patient level testing, 50 documents from annotation level testing, search term 'hypertension*', 'high blood pressure*')

Search Terms

Ontology available on request

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1	The application excludes the following phrases: 'possibility [0-5 words] hypertension' 'risk of [0-8 words] hypertension' 'possibl* hypertension'	Random sample of 200	P=94%			

	'suspected hypertension' 'formcheckbox hypertension' And FORMS 1, 4, 5, 6, 7 and 8 (details available on request)					
2	As above See post-processing rules in Notes			Random sample of 100 for precision and 50 for recall	Patient-level P=94% R=94%	hypertension* high blood pressure*

NOTES

Remaining false positives refer to: side effects/risk of hypertension and a couple of family hypertension mentions.

No pattern could be seen in the false negatives raised.

Post-processing rules

Where phrase_exclude = 0

Where (form_exclude = 0 or (form_exclude = 1 and form_exclude_type like 'form_3%') or (form_exclude = 1 and form_exclude_type like 'form_2%') or (form_exclude = 1 and form_exclude_type like 'form_9%') or (form_exclude = 1 and form_exclude_type like 'form_10%'))

We found common forms and phrases within patient records that reduce precision and were thus excluded from the application. The list of forms is available on request.

Production

- Run schedule – monthly

Version – 1

8. MULTIMORBIDITY – 21 LONG-TERM CONDITIONS (MEDCAT)

Description

Application to identify patients with diagnosis of physical health conditions (21 conditions in total, including hypertension, diabetes, epilepsy etc.)

Definition

Development approach: machine learning

Classification of past or present symptom: Both.

Classes produced: Positive.

Positive mentions include:

'He reported that he suffers from diabetes and hypertension', 'Ms zzz has a history of atopy including asthma, 'physical health history: asthma, diabetes, high blood pressure'; 'Nil other problems other than epilepsy'; 'Physical health: lung disease confirmed'

Interrater reliability

N/A

Search Terms

5 most commonly searched terms were used for each condition

Performance

Conditions	Annotated documents identified by the application	Precision	Un-annotated documents extracted from keyword search in CRIS	Recall	Keywords used for extraction from CRIS
Cerebrovascular accident	N=156	96%	N=156	89%	stroke CVA Stroke strokes CVAs
Epilepsy	N=214	95%	N=214	84%	epileptic epilepsy Epilepsy epileptic fits epileptic seizures

Diabetes mellitus	N=217	90%	N=217	84%	diabetes diabetic Diabetes Diabetic diabetes mellitus
Chronic kidney disease	N=175	99%	N=175	99%	CKD ckd chronic kidney disease Chronic kidney disease Chronic Kidney Disease
Psoriasis	N=135	97%	N=135	80%	psoriasis Psoriasis psoariasis psoraisis psorasis
Parkinson’s disease	N=143	97%	N=143	50%	Parkinsons disease Parkinsons Parkinson’s Parkinson’s disease Parkinsons
Multiple sclerosis	N=141	93%	N=141	85%	multiple sclerosis Multiple Sclerosis Multiple sclerosis MULTIPLE SCLEROSIS multiple sclerosis (MS
Eczema	N=123	92%	N=123	87%	eczema Eczema dermatitis ezcema Dermatitis
Hypertensive disorder, systemic arterial	N=111	92%	N=111	91%	hypertension Hypertension hypertensive

					high blood pressure HTN
Transient ischemic attack	N=124	98%	N=124	95%	TIA TIAs Transient ischaemic attack transient ischaemic attack transient Ischemic Attacks (ITAs
Migraine	N=125	93%	N=125	81%	migraine migraines Migraine a migraine Migraines
Chronic obstructive lung disease	N=126	94%	N=126	84%	COPD Chronic Obstructive Pulmonary Disease chronic obstructive pulmonary disease Chronic obstructive pulmonary chronic obstructive airways disease
Arthritis	N=122	99%	N=122	81%	arthritis Arthritis Rheumatoid arthritis rheumatoid arthritis Rheumatoid Arthritis
Heart failure	N=117	92%	N=117	84%	heart failure Heart failure cardiac failure Heart Failure CCF
Asthma	N=124	98%	N=124	78%	Asthma asthma

					asthmatic Asthmatic Asthama
Ischemic heart disease	N=125	92%	N=125	97%	heart attack angina IHD myocardial infarction Angina
Inflammatory bowel disease	N=112	99%	N=112	90%	ulcerative colitis inflammatory bowel disease Ulcerative Colitis Ulcerative colitis IBD
Atrial fibrillation	N=115	99%	N=115	91%	atrial fibrillation Atrial fibrillation Atrial Fibrillation Atrial Fibrillation (AF atrial fibrilation
Chronic liver disease	N=109	99%	N=109	78%	chronic liver disease Chronic Hepatitis chronic hepatitis Chronic hepatitis chronic Hepatitis
Chronic sinusitis	N=105	97%	N=105	93%	chronic sinusitis Chronic Sinusitis Chronic sinusitis Chronic Rhinasinusitis Chronic rhinosinusitis
Coronary arteriosclerosis	N=103	97%	N=103	86%	Coronary artery disease

NOTES

Please note, use of these applications will need prior discussion and approval with the CRIS team while they are being implemented and under further evaluation.

Post-processing rules

N/A

Production

- Run schedule – monthly

Version – 1

9. PAIN

Description

The purpose of this application is to determine if a mention of pain (or related words, such as sore, ache, *algia, *dynia etc.) within the text is relevant i.e. associated with the patient and refers to physical pain.

Definition

Development approach: Machine learning

Classification of past or present symptom: Both.

Classes produced: Positive.

Positive mentions include:

'she is in constant pain'; 'he suffers from severe headaches'; he is taking pain killers due to a pulled muscle'

Interrater reliability

Cohen's k = 86% for Attachment (based on 865 annotations)

Cohen's k = 91% for Event (based on 458 annotations)

Search Terms

%dynia%, '%algia%, %burn%', % headache%, % backache%, % toothache%, % earache%, % ache%, %sore%, %spasm%, % colic%, % cramp%, % hurt%, % sciatic%, % tender%, % pain %, % pains%, % painful%

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 (1 AddictionEvent, 23 Attachment, 3 CAMHS Event, 1 CCS_Correspondence, 2 Discharge_Notification_Summary, 52 Event, 2 Mental_state_formulation, 1 POSProforma, 2 RiskAssessmentTool_RiskFactors, 1 Single_generic_Assessment, 11 Ward Progress Note, 1 WardRound	P=91%	Random sample of 100 (50 Attachment, 50 Events)	R=78%	%dynia% %algia% %burn% % headache% % backache% % toothache% % earache%

						% ache% %sore% %spasm% % colic% % cramp% % hurt% % sciatic% % tender% % pain % % pains% % painful%

NOTES

N/A

Production

- Run schedule – monthly

Version – 1

10. RHEUMATOID ARTHRITIS

Description

Application to identify patients with diagnoses of rheumatoid arthritis.

Definition

Development approach: sem-EHR**.**

Classification of past or present symptom: Both.

Classes produced: Positive.

Positive mentions include:

'ZZZZZ has been in pain due to her rheumatoid arthritis', 'she has been bedbound with rheumatoid arthritis this week', 'medication for her rheumatoid arthritis', 'physical health comorbidities: hypertension, rheumatoid arthritis', 'diagnosed with rheumatoid arthritis is 1988'

Interrater reliability

Cohen's k = 98% (50 patients from patient level testing, 50 documents from annotation level testing, search term 'rheumatoid arthritis')

Search Terms

Ontology available on request

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1	The application excludes the following phrases: 'possibl* rheumatoid arthritis' 'suspected rheumatoid arthritis'			Random sample of 100 for precision and 50 for recall	Patient-level R=91% R=86%	rheumatoid arthritis

	‘formcheckbox rheumatoid arthritis’ ‘possibility [0-5 words] rheumatoid arthritis’ ‘rheumatoid arthritis and other inflammatory arthropathy’ ‘?rheumatoid arthritis’ See post-processing rules in Notes					

NOTES

Remaining false positives occur due to:

1. Uncertain diagnoses:

‘has she been diagnosed with rheumatoid arthritis?’

‘(still waiting for the results, including enquiry re Rheumatoid Arthritis)’

‘differential diagnosis was ankylosing spondylitis; rheumatoid arthritis’

2. Undetected forms/headings:

‘Antibodies to citrullinated peptide or citrullinated filaggrin are highly specific for Rheumatoid Arthritis’

‘Rheumatoid arthritis, Hashimoto's thyroiditis and lupus are examples of autoimmune diseases’

‘The physical symptoms can be as disabling as multiple sclerosis, systemic lupus, rheumatoid arthritis and other chronic conditions (NICE guidelines 2007)

There was no pattern of false negatives, which included failures such as:

‘diagnosed with rheumatoid arthritis'

'he said he has rheumatoid arthritis'

'she also has a history of rheumatoid arthritis’

Post-processing rules

Through testing, we have found that optimum precision and recall are indicated when documents containing >2 mentions of the illness are not excluded. We do not exclude documents containing the above exclusion terms if they contain >2 mentions of the illness.

Production

- Run schedule – monthly

Version - 1

11. HIV

Description

Application to identify instances of HIV diagnosis.

Definition

Development approach: Machine learning**.**

Classification of past or present symptom: Both.

Classes produced: positive, negative and unknown.

Include cases where a definite HIV diagnosis is present in the text (e.g. ZZZZZ was diagnosed with HIV)

Exclude all other cases

Interrater reliability

Search Terms

hiv

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 (1 addiction physical health assessment, 2 addictions events, 32 attachments, 1 care plan mental health, 1 care plan physical health, 1 ccs correspondence, 5 discharge notifications, 39 events, 1 Risk Assessmen tTool_CRISRiskPlan, 1 Risk Assessment Tool_RiskFactors, 2 single generic assessments, 2 summaries of	P=94%	Random sample of 100 (50 events and 50 correspondence)	P=70% P=100%	hiv

		need, 12 ward progress notes)				
2		Random 50 patients	P=92% Patient-level P=64%			

Production

- Run schedule – on request
- Version – 1

12. HIV TREATMENT

Description

Application to identify instances of HIV treatment.

Definition

Development approach: Machine learning***.***

Classification of past or present symptom: Both.

Classes produced: positive, negative and unknown.

Include any positive references to the search terms below.

Exclude all other cases.

Interrater reliability

Search Terms

Anti-retroviral
antiretroviral
ARV
HAART
cART
ART
CD4
Undetectable
Abacavir
Lamivudine
Zidovudine
Aptivus
Atazanavir
Atripla
Celsentri
Cobicistat
Combivir
Darunavir
Didanosine
Dolutegravir
Edurant
Efavirenz
Elvitegravir
Emtricitabine
Emtricitabine
Emtriva
Enfuvirtide
Epivir
Etravirine
Eviplera
Fosamprenavir

Fuzeon
Indinavir
Intelence
Invirase
Isentress
Kaletra
Kivexa
Lamivudine
Zidovudine
Lopinavir
Ritonavir
Saquinavir
Stavudine
Stribild
Sustiva
Telzir
Tenofovir
Efavirenz
Emtricitabine
Emtricitabine
Elvitegravir
Cobicistat
Tenofovir
Emtricitabine
Rilpivirine
Tipranavir
Tivicay
Trizivir
Truvada
Tybost
Videx
Viramune
Viread
Vitekta
Zerit
Ziagen
Zidovudine
Co-trimoxazole
Cotrimoxazole
Septrin

Performance

	Post-processing rules added	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from	Precision and recall (un-annotated)	Keywords used for extraction from CRIS

	to application			**keyword search in CRIS**		
1		Random sample of 100 (1 addiction physical health assessment, 37 attachments, 1 ccs correspondence, 4 discharge notifications, 35 events, 1 Risk Assessment Tool_RiskFactors, 1 risk event, 2 triage forms ARC, 15 ward progress notes, 3 ward rounds)	P=94%	Random sample of 100 (50 events and 50 correspondence)	P=98% P=100%	viral load antiretroviral* ritonavir truvada tenofovir
2		Random 50 patients	P=95% Patient-level P=76%			

Production

- Run schedule – on request
- Version – 1

CONTEXTUAL FACTORS

1. AMPHETAMINE

Description

To identify instances of amphetamine use.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include “denies current use of amphetamine, however last reported using 3 months ago”, “first took amphetamines at the age of 15”, “UDS: +ve amphetamine”, “ZZZZZ has been trying to give up amphetamine for the last 2 months”, “ZZZZZ was found in possession of large quantities of amphetamines”, “She admitted to having bought amphetamine 2 days ago” , “amphetamine-psychosis”

NB. Assumption that if bought cocaine/crack then has also taken it. This is subjective and should be decided by the annotator. It is more important that the annotator is consistent than “right” about classifying this sentence. Even though “stopped” or “gave up” suggest a present lack of exposure, they also indicate a past use and therefore are classified as positive.

Negative annotations include “ZZZZZ denies use of alcohol and amphetamine”, “ZZZZZ has not used amphetamine for the last week”, “-ve: amphetamine”

N.B. Although an addition like “since yesterday” to the negation may suggest that cocaine was taken previously, we still classified a negation as negative.

Unknown annotations include “ZZZZZZ’s mother has a history of amphetamine abuse” – subject other than patient, “ZZZZZ is planning on taking amphetamine this weekend” – future or conditional event, “We discussed the dangers of amphetamine”

Interrater reliability

Cohen's k = 84% (50 un-annotated documents - 25 events/25 attachments, search term ‘amphetamine*’)

Search Term

amphetamines-have

amphetaminergic

amphetamines-makes

Amphetamine-prescribed

amphetamine-induced

Amphetamine---

amphetamine-induce

amphetamine-based

Amphetamine-

amphetamines-

amphetamineStarted

Amphetamine

Amphetamines

amphetamine-type

amphetamine-sulphate

amphetamines-using

amphetamine-driven

amphetamine-like

amphetamine-family

amphetamine-which

Amphetamine-related

amphetamines-paranoia

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 30 (one document per patient)	P=90%			
2		Random sample of 100 - 6 drug and alcohol history, 51 event clinical notes, 43 correspondence- attached text	P=76%	Random sample of 100 – 50 attachments, 50 events	P=80% R=84%	amphetamine*

NOTES

All false positives were found in correspondence- attached text comments. 6 were classed as negatives (negations: e.g. never taken, not used amphetamines). The rest were classed as unknowns, all having the mention of amphetamine within a list to be ticked if patient has/has not been exposed to the substance. An example being: '*FORMCHECKBOX Amphetamines FOMRCHECKBOX Other (please specify) 7.3'.*

While current and past use were both labelled as positive, I also labelled whether each positive mention was describing past or present exposure. I categorised past exposure as history of use, describing one specific past incident, or mentioning regular use with emphasis on the patient having stopped now. I categorised present exposure as current use, addiction of, a positive urine test and mention of a regular incident pattern eg uses 2x weekly. The majority of positive mentions were present use (63.2%) compared to past use (36.8%).

There was a contradiction between positive/negative instances. Mentioning having 'stopped' was labelled as a positive (as it references past use), however stating 'has not used' in past week would be labelled as a negative, despite them both meaning the same thing. This also means that those who have never used and those who have used in the past are both classified as negative, due to a negation term being used.

There was only one positive instance where being exposed to crack was classed as a positive.

Production

- Run schedule - monthly
- Version - 1

2. CANNABIS

Description

To identify instances of cannabis use.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include he is a cannabis smoker, she smoked cannabis when at uni. Include cases where there is a reference to stopping use but not explicit reference to current use e.g., she stopped using cannabis 3 years ago.

Negatives annotations include denied taking any drugs including cannabis, no cannabis use.

Unknown annotations include she stated in hash voice, pot of yoghurt, father cannabis user, pot for UDS.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search terms ‘cannabis ‘, ‘marijuana’, ‘weed’, ‘pot’, ‘hash’, ‘skunk’, ‘resin’, ‘spice*’)

Search Terms (case insensitive)

cannabis

skunk

weed

Pot

marijuana

grass

THC

hash

cannabinoids

resin

hashish

weeds

Cannabis-

spices

Spice

ganja

CBD

cannabis-induced

Cannabinoid

cannabies

grasses

Cannaboids

marijuanna

cannabbase

cannabis-free

skunk-

cannabbis

Hashis

cannabis-related

cannabi

cannabise

cannabinoides

cannabis-use

marijuna

cannabus

cannabiss

weed-

skunks

Cannabises

cannabis--

cannaboid

cannabid

THC-

pro-cannabis

cannabinoids-

cannabanoids

cannabsi

cannabls

use-cannabis

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 30 (one document per patient)	Overall P=93% Without spice/ cannabinoid/cannaboid P=93% cannabinoid/spice only P=74% Negative P=48%			
2		Random sample of 100 - 20 correspondence-attached text, 1 mental health care plan, 6 discharge brief summaries, 2 drug and alcohol histories, 2 ward	Overall P=88% Current P=72%	Random sample of 100 – 50 attachments, 50 events	Overall P=80% R=88% Current P=59% R=86%	cannabis marijuana weed pot hash skunk resin spice*

		progress notes, 7 mental state assessment summaries, 62 event clinical notes				
3	Application excludes instances of ‘*cannabinoid*’, ‘*cannaboid*’ or ‘*spice*’ (see notes)	Random sample of 100 - 20 corresponde nce-attached text, 1 mental health care plan, 6 discharge brief summaries, 2 drug and alcohol histories, 2 ward progress notes, 7 mental state assessment summaries, 62 event clinical notes	Overall P=88% Current P=72%	Random sample of 100 – 50 attachments , 50 events	P=77% R=93%	cannabis marijuana weed pot hash skunk resin

NOTES

False positives were mainly references when the term ‘pot’ was irrelevant e.g. pot of yogurt or pot for urine testing. Often many references to cannabis use were consistently flagged in the same document.

Code for post-processing

Name not like ‘%cannabinoid%’, ‘%cannaboid%’ or ‘%spice%’

Production

- Run schedule – monthly
- Version - 1

3. COCAINE OR CRACK COCAINE

Description

To identify instances of cocaine or crack cocaine use.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include “denies current use of cocaine, however last reported using 3 months ago”, “first smoked cocaine at the age of 15”, “UDS: +ve cocaine”, “ZZZZZ has been trying to give up cocaine for the last 2 months”, “ZZZZZ was found in possession of large quantities of cocaine”, “She admitted to having bought cocaine 2 days ago” , “He has stopped taking cocaine”.

N.B. Assumption that if bought cocaine/crack then has also taken it. This is subjective and should be decided by the annotator. It is more important that the annotator is consistent than “right” about classifying this sentence. Even though “stopped” or “gave up” suggest a present lack of exposure, they also indicate a past use and therefore are classified as positive.

Negative annotations include “ZZZZZ denies use of street drugs such as cocaine”,“ZZZZZ has not used cocaine for the last week”, “Crack N” – form style.

N.B. Although an addition like “since yesterday” to the negation may suggest that cocaine was taken previously, we still classified a negation as negative.

Unknown annotations include “ZZZZZZ’s mother has a history of crack abuse” – another subject other than the patient, “ZZZZ is planning on taking cocaine this weekend” – future or conditional events, “When cooking he decided to crack the eggs open” – irrelevant , “ZZZZZ believes cocaine isn’t good for people” – irrelevant, “We discussed the dangers of crack”.

Also include statements such as ‘He did not smoke cocaine today’- unclear whether past use or never use.

Interrater reliability

Cohen's k = 95% (50 un-annotated documents - 25 events/25 attachments, search term ‘cocaine*’)

Search Terms

cocaine

Cocaine-

COCAINE--

Cocaine----

cocaine--this

cocaine-based

cocaine-cannot

cocaine-cautioned

cocaine-dealing

cocaine-dependence

cocaine-ecstacsy-has

cocaine-for

cocaine-greatly

cocaine-he

cocaine-however

ocaine-induced

cocaine-initially

Cocaine-it

cocaine-laced

cocaine-last

Cocaine-lasted

cocaine-managed

cocaine-most

cocaine-not

Cocaine-occasional

cocaine-positive

cocaine-postitive

cocaine-presented

cocaine-referred

cocaine-related

cocaine-smoking

Cocaine-snorting

cocaine-some

Cocaine-started

cocaine-surely

cocaine-trigger

cocaine-up

cocaine-use

Cocaine-used

Cocaine-uses

Cocaine-using

cocaine-was

cocaine-weekend

cocaineamytriptilline

cocaineapprox

cocaineat

cocained

cocainefor

cocaineher

cocainehowever

cocainel

cocainein

cocaineingestion

cocainelast

cocainemetabolite

cocaineon

cocainer

cocaines

Cocainestarted

cocainetubes

cocaineuse

cocainex

Crack

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 30 (one document per patient)	Overall P=97%			
2		Random sample of	P=79%	Random sample of	P=84%	crack*

		100 - 70 event clinical notes, 3 CCS correspondence texts, 1 mental health care plan, 26 correspondence-attachment text		100 – 50 attachments, 50 events	R=97%	*cocaine*

NOTES

False positives occurred when mentions were of patients describing other individuals as crack users or describing what those users are like in general. False positives also occurred when mentions were of an individual that was not the patient. Unknowns were questions of patient's use of cocaine/crack, vague comments e.g. 'appears to be', and when the patient dealt the drug to other individuals for profit.

While current and past use were both labelled as positive, I also labelled whether each positive mention was describing past or present exposure. I categorised past exposure as history of use, describing one specific past incident, or mentioning regular use with emphasis on the patient having stopped now. I categorised present exposure as current use, addiction of, a positive urine test and mention of a regular incident pattern e.g. uses 2x weekly. The majority of positive mentions was past use (62%) compared to present use (38%).

There was a contradiction between positive/negative instances. Mentioning having 'stopped' was labelled as a positive (as it references past use), however stating 'has not used' in past week would be labelled as a negative, despite them both meaning the same thing. This also means that those who have never used and those who have used in the past are both classified as negative, due to a negation term being used.

In all cases, 'crack cocaine' was classed as two positive instances (crack and cocaine independently).

Production

- Run schedule - monthly
- Version - 1

4. MDMA

Description

Application to identify instances of MDMA use.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include "denies current use of MDMA, however last reported using 3 months ago","first took MDMA at the age of 15", "UDS: +ve MDMA", "ZZZZZ has been trying to give up MDMA for the last 2 months", "ZZZZZ was found in possession of large quantities of MDMA", "She admitted to having bought MDMA 2 days ago" . "He has stopped taking MDMA".

N.N. Assumption that if bought MDMA then has also taken it. This is subjective and should be decided by the annotator. It is more important that the annotator is consistent than "right" about classifying this sentence. Even though "stopped" or "gave up" suggest a present lack of exposure, they also indicate a past use and therefore are classified as positive

Negative annotations include "ZZZZZ denies use of street drugs such as MDMA" ,"ZZZZZ has not used MDMA for the last week", "UDS -ve: MDMA".

N.B. Although an addition like "since yesterday" to the negation may suggest that MDMA was taken previously, we still classified a negation as negative.

Unknown annotations include "ZZZZZZ's mother has a history of MDMA abuse" – another subject other than the patient, "ZZZZ is planning on taking MDMA this weekend" – future or conditional events, "ZZZZZ believes MDMA isn't good for people" – irrelevant, "We discussed the dangers of MDMA", "MDMA".

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'mdma')

Search Terms (case insensitive)

mdma

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 30 (one	P=87%			

		document per patient)				
2		Random sample of 100 - 7 ward progress notes, 10 CCS correspondence - attached text, 1 CAMHS event, 22 correspondence - attached text, 60 event-clinical notes	P=94%	Random sample of 100 – 50 attachments , 50 events	P=100% R=99%	mdma

NOTES

False positives occurred when there was suspected MDMA use, future planned use that hadn't been undertaken yet and one negation 'denies' use. The couple of unknown mentions were when MDMA was part of a list without direction as to whether MDMA use was prevalent.

Production

- Run schedule - monthly
- Version - 1

5. SMOKING

Description

This application distinguishes between people who are a) current smokers, b) current non-smokers (ever smoked) and c) non-smokers. This application may at times bring back contradictory information on the same patient since patient may start smoking and stop smoking and because of the varied level of information available to the clinician.

Definition

Development approach: Rule-based***.***

Annotation Rules

Status:

One of the following must be annotated in the status feature:

Never = clearly not smoking currently or just a general message that the subject does NOT smoke. Ex: "…is a non-smoker", "… was/is not a smoker", "… doesn't smoke", "ZZZZZ denies ever smoking", or "… is currently not smoking"

Current = a clear message that the subject is currently smoking

Ex: "…smokes 20 cigarettes a day", "… has been smoking for 10 years", "…is a smoker", "ZZZZZ smokes in the ward", "…went to garden for a smoke", "ZZZZZ is stable when smoking", "…has a history of heavy smoking", "Consider stopping smoking", "ZZZZZ found smoking in her room" or "… is a tobacco user")

Past = any hint that the subjects was smoking

Ex: "… used to smoke", "… has quitted smoking", "… stopped smoking", "ZZZZZ is an ex-smoker" or "…was a smoker")

Subject:

One of the following must be annotated in the subject feature: "patient" or "other". For the most cases, the information of smoking is about the subject him/herself. But, there is still a need to exclude the "noise" from "other" smokers. If there is no subject in the whole sentence, it should be considered as the subject is the patient him/herself. ZZZZZ is the symbol used for anonymising patient's name. QQQQQ is now used to anonymise someone other than the subject and staff in clinics or hospitals. If no clear information could be identified for subject feature within the whole sentence (ex., "He stopped smoking for years"), the subject should be taken as the patient.

Examples:

Advised by GP for smoking cessation – "current" and "patient"

Bought tobacco – "current" and "patient"

Used the smoking room – "current" and "patient"

has stopped smoking for years – "past" and "patient"

;;;;; Smoking;;;;; - "current" and "patient"

…doesn't smoke – "never" and "patient"

…is quitting smoking – "current" and "patient

…stopped smoking for 2 years – “past” and “patient”

N.B. This app may at times bring back contradictory information on the same patient since patients may start smoking and stop smoking and the level of information available to the clinician may vary.

Interrater reliability

N/A

Search terms

N/A

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1				Random sample of 100 – 50 attachments, 50 events	Smoking mention All documents P=85% R=89% Events only P=97% R=88% Attachments only P=77% R=89%	*smok* *cigar* *tobacco*
2				All positive hits from above sample	Smoking status Current P=79% R=87% Past	*smok* *cigar* *tobacco*

					P=68% R=38% Never P=72% R=75%	
3		Random sample of 100 - 7 physical health, 18 mental health, 1 drug and alcohol history, 1 assessment-presenting circumstances and 53 event clinical notes	Overall P=92% Status P=97% Subject P=35%	Random sample of 100 – 50 attachments, 50 events	Overall P=81% R=74%	*smok* *cigar* *tobacco*
4		Random sample of 40 for each category – overall, current, never, past	Overall P=83% Current P=90% Never P=73% Past P=55%			
5		Random sample of 90 – 28 attachments, 1 MH care plan, 1 PH care plan, 45	P=86% Patient- level P=95%			

		events, 14 ward progress notes				
6		Random sample of 100 with F2* diagnosis – 57 attachments, 43 events	P=81% Patient- level P=94%			
7		Random sample of 60 with F2* diagnosis and 1 annotation per patient – 26 attachments, 3 MH care plans, 7 PH care plans, 3 ccs corresponden ce, 2 discharge notification summaries, 19 events	P=55% Patient- level P= 79%			
8		Random sample of 100 with F2* diagnosis and latest annotation per patient – 30 attachments, 1 CAMHS event, 2 ccs corresponden ce, 2 discharge notification summaries, 19 events	P=60% Patient- level P=75%			

9		Random sample of 93 documents, 1 annotation per patient for those with 10+ annotations	P=96% Patient- level P=96% R=62%			
10		Random sample of 118 documents, 1 annotation per patient for those with 5+ annotations	P=90% Patient- level P=90% R=78%			

NOTES

False positives occurred when irrelevant comments were made relating to smoke from a fire, smoke alarm function or fire alarm procedure. False positives also occurred when hypothetical 'if' situations were used. Comments were classed as unknown if referring to smoking cannabis (that may contain some tobacco), while smoking heroin (would not contain tobacco) was labelled as a negative mention.

The precision of status was very good, with only three instances of incorrect labelling: labelling as current instead of never (x2) and never instead of current (x1).

The precision of 'who' mainly occurred when the app classed a mention as none/NULL instead of patient, suggesting an inability of the app to identify when the note is referring to the patient.

Sometimes the app was able to identify the patient in some instances but not others within the same document. Many of the cases where patient was not identified was relating to patient's access to the smoking room, talking about smoking cessation services (not yet attended or ineffectiveness of them). However, a few were also direct smoking mentions that were not detected.

When applying the smoking application to a population with F2* diagnoses, the best performance is achieved by using patients with >5 'current' annotations.

Production

- Run schedule – weekly
- Version - 1

6. ONLINE ACTIVITY

Description

Application to identify and distinguish between mentions of internet/social media/online gaming in patient records across Child and Adolescent Mental Health Services.

Definition

Development approach: Rule-based***.***

Classifications: INTERNET, ONLINE_GAMING, SOCIAL_MEDIA.

Internet

We are interested in patterns and the nature of internet use and content viewed online. Online platforms such as *Pinterest, YouTube* or specific websites may be documented. In some cases, there is insufficient detail to establish what online activity is being engaged with *i.e., "… spends a lot of time online"*. In these cases, and where the mention is clearly related to online activity, it should be annotated as 'Internet'.

Social Media

Social media is defined as *websites and applications that enable users to create and share content or to participate in social networking.* Mentions may refer to specific platforms included in the gazetteer such as: *Instagram, Twitter, Facebook, Snapchat*, or to a behaviour i.e. *"Chatting to their friends online".*

Online Gaming

We are interested in online gaming and have included general terms and more specific titles of games such as *Call of Duty, Fortnite, Minecraft*. Games consoles i.e. *Playstation, Xbox and Nintendo DS* have also been included in the gazetteer as they increasingly have enhanced online functions. Some online gaming mentions will be less specific and refer to behaviour, for example: *"Spends a lot of time playing video games", "likes playing games on the internet with her friends"* but should still be coded.

'Other' online use

In view of the fact that social media and internet activity are often accessed via mobile devices we have also included: *iPhone, iPad, Blackberry, Smartphone.* Where there is suggestion that these are used for online gaming or social media they should be annotated accordingly. If the exact use is not clear they are annotated as INTERNET.

Interrater reliability

Number of matching files: 149

Inter-rater agreement (test)		
	Spans	**Attributes**
Precision (macro)	0.9	0.97
Recall (macro)	0.82	0.95
F-score (macro)	0.86	0.96
Precision (micro)	N/A	0.97
Recall (micro)	N/A	0.97
F-score (micro)	N/A	0.97

Kappa	N/A	0.94

Search Terms (case insensitive)

Gazetteer available on request

Performance

A test corpus (n=6172) was randomly divided between two researchers (human inter-rater agreement 0.94) and all relevant mentions of online activity were annotated according to the annotation guidelines.

Number of documents (annotated and unannotated): 6,172

Number of annotations: 535

Evaluation results (test)		
	Spans	**Attributes**
Precision (macro)	0.73	0.97
Recall (macro)	0.76	0.94
F-score (macro)	0.74	0.95
Precision (micro)	N/A	0.95
Recall (micro)	N/A	0.95
F-score (micro)	N/A	0.95
Kappa	N/A	0.92

NOTES:

Most common false positive is insufficient contextual disambiguation for the following words: computer, Internet, mobile phone, online, PC, website. It performed less well distinguishing class from longer spans of free text i.e.

Gold: 1156 1189 playing games with friends online ...

System: 1183 1189 online

-- attribute disagreement on class: ONLINE_GAMING vs. INTERNET

MISSING ANNOTATIONS

9266 9298 playing games a lot on his phone

MATCHING ANNOTATIONS

Gold: 618 639 games on the computer

System: 631 639 computer

-- attribute disagreement on class: ONLINE_GAMING vs. INTERNET

Mention of all specific websites described in CRIS would not be feasible, but inclusion of www. …. co.uk or other more generic identifiers resulted in too many false positives (i.e. the NHS Trust or affiliated websites contained in letter headers). Similarly, ‘email*’ generated too many false positives during development to be included. These may therefore be false negatives that should be considered when using the NLP application.

Production

- Run schedule – monthly
- Version - 1

7. EDUCATION

Description

Application to identify the highest level of education at patient level.

Definition

Development approach: Rule-based***.***

The Education application will produce 4 features for each annotation:

Group: A levels/GCSE/University/unqualified

Subject: patient/uncertain

Rule: Annotations for each group will be assigned independently of each other, e.g. in theory the same text could produce annotations in each group.

The Education application can also be used to extract information about school leaving age.

Group 1: A level group

Rule	**Stage of course**
Accepted	Accepted for A-level course or equivalent (course or institution)
Ongoing	Started course but not (yet) completed (including evidence of attending relevant institution)
Dropped out	Started course but not completed - dropped out
Expelled	Started course but not completed - expelled
Failed	Completed course – failed all exams
Completed	Completed course
Passed	Passed at least one exam
Applied_undergrad	**Applied for university / course**

Note: aspirations, plans, application only are not accepted.

Group 2: GCSE group

Rule	**Stage of course**
Ongoing	Started GCSE course (or equivalent) but not (yet) completed
Completed	Completed GCSE course or equivalent
Passed	Passed at least one exam (GSCE or equivalent)
Applied_A-level	**Applied for 6th form (college) / A-level**

Group 3: University

Rule	Stage of course
Accepted	Accepted for course / institution
Ongoing	Started course but not (yet) completed
Dropped out	Started course but not completed - dropped out
Expelled	Started course but not completed - expelled
Failed	Completed course – failed
Completed	Completed course
Passed	Passed / graduated
Applied_University	**Applied for University**

Group 4: unqualified group

Rule	Definition
Unqualified	A specific reference in notes describing as having left school without any qualifications.
GSCE_Dropped_out	**Started GCSE course but not completed - dropped out**
GSCE_Expelled	**Started GCSE course but not completed - expelled**
GSCE_Failed	**Completed GCSE course – failed all exams**

School leaving age

Examples
He left school at the age of 16 years
Was 19 years old when she left school
Mrs ZZZZZ left school at 15 without any qualifications

Interrater reliability

GCSE – Cohen's k = 90% (50 annotated documents – 25 events, 25 attachments)

No qualifications - Cohen’s k = 100% (50 annotated documents – 25 events, 25 attachments)

Search Terms (case insensitive)

Gazetteer available on request

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 - 100 CAMHS-clinical notes	General P=70% Level P= 99% Who P= 75.7%	Random sample of 100 - events and attachments	P=86% R=23%	a-level* cse* a level*
2		Random sample of 100 - 2 ward progress notes, 1 social situation, 1 mental health formulation, 1 presenting circumstances, 4 personal histories, 1 discharge summary, 1 mental health care plan, 4 CCS correspondence, 7 CAMHS event-clinical notes, 49 correspondence-attached	General P=89% Who P=89%	Random sample of 100 - events and attachments	P=91% R=23%	gcse*

		text, 29 events-clinical notes				
3		Random sample of 100 - 52 attachments, 3 camhs events, 3 ccs correspondence, 2 discharge notification summaries, 30 events, 4 personal histories, 1 presenting circumstances, 4 ward progress notes	General P=87% Subject P=81% Patient-only P=85%	Random sample of 100 - events and attachments	General P=91% Recall=84%	6^{th} form college AS in AS level examinations AS level exams Diploma NVQ 3 National diploma A level A level To college to study
4		Random sample of 100 - 53 correspondence-attached text, 1 CAMHS event-clinical note, 1 discharge summary, 1 CCS correspondence-attached text, 3 ward progress notes, 1 presenting circumstance, 7 personal histories, 40	General P=94% Who P=83%	Random sample of 100 - events and attachments	P=77% R=15%	no gcse* without gcse* failed gcse* incomplete gcse* without any gcse* without any qualification* Without qualification* without formal qualification* without any formal qualification*

		events- clinical notes				
5				Random sample of 100 – personal history	Patient-level P=55%	*a-level* *diploma* *o level * gcse* college* exams* sixth form* school* uni* graduate* without any qualification* without qualification* without formal qualification* without any formal qualification* university* *degree* *phd* *masters* * MA * *city and guilds* * BSc * NB. School* was removed after 37 annotations
6				Random sample of 60 – events and attachments	Patient-level P=65%	As above
7	Application excludes records of patients < 18 years old	Random sample of 100	Patient-level P=83%			As above
8	Application excludes records of	Random sample of 100 – 50	Patient-level P=80%			As above

	patients < 18 years old	events, 50 attachments	R=81%			
9		Random sample of 100 – 48 attachments, 2 CAMHS events, 7 CCS correspondence, 27 events, 15 histories, 1 social situation	P=100%	Random sample of 100 – 50 events and 50 attachments	P=100% R=51%	*left school*
10		Random sample of 100 – 44 attachments, 4 CCS correspondence, 1 discharge notification summary, 22 events, 6 histories, 1 mental state formulation, 1 risk assessment too Risk Factors, 14 single generic assessments, 3 social situations, 3 summaries of need, 1 ward progress note	P=98%	Random sample of 100 – 50 events and 50 attachments	P=98% R = 76%	*left school*

NOTES

Round 5

There was no seen pattern in false level instances. It was also unsure whether low precision was due to the app or due to personal histories not encompassing the general education level (present in other documents).

Round 6

Over half of the documents were NULL, due to CAMHS involvement: children were too young to have a qualification. Most of the errors were in these documents, as children aspired to go to university/ college (labelled as a positive instance by the app).

Round 7

Most false positives were due to not recognising the GNVQ – level 1.2 qualification, mention of MA (labelled as university falsely) and hypothetical mention of applying to university.

Rounds 7 and 8

75% of false positives were due to the app labelling education level as university when it was actually lower (gcse or a level). These were often due to the 'MA' abbreviation being misunderstood, hypothetical mentions of applying for university or thinking of applying or irrelevant mentions of someone else going to university (eg. Child/sister). Other false positives were due to problems with the GNVQ qualification (usually classed at a higher level than it is). This might be hard for the app to distinguish as GNVQ level 1 and 2 could be GCSE or A level. False negatives were usually due to the NVQ qualification, classed as null rather than a gcse level (level 2). Other false negatives were mentions of leaving school when the mention did not have the word 'qualification' in it. eg. 'left school at 14', 'left school without Q's' 'limited schooling' and 'no formal education after age of 13' were classified as null.

Production

- Run schedule – on request
- Version - 1

8. OCCUPATION

Description

Application to identify occupations/work descriptions and who these refer to.

Definition

Development approach: machine-learning and rules-based***.***

Classification of past or present occupation: Both.

There are two parts to each annotation: Firstly, the **occupation feature** is annotated - this could be a job title, for example a 'builder'; or a job description, for example 'working in construction'. Secondly, the **occupation relation** is annotated: who the occupation belongs to, for example the patient or their family member.

Unpaid occupational categories were included (e.g. student, unemployed, homemaker, volunteer). Depending on the text available, extractions can state a specific job title (e.g. head-teacher) or a general occupational category (e.g. self-employed).

Work aspirations were excluded from annotations. Frequently extracted health/social care occupations (e.g. psychiatrist) are not annotated as belonging to the patient, in order to maximise precision.

Occupation feature (text) – the job title (e.g. 'hairdresser')

Occupation relation (text) – who the occupation belongs to (e.g. 'patient')

The full annotation guideline document is available on request.

Interrater reliability

Cohen's k = 77% for occupation feature (200 'personal history' documents)

Cohen's k = 72% for occupation relation (200 'personal history' documents)

Search Terms

Gazetteer available on request

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1	The application employs a filter by which healthcare-related occupations such as psychiatrist,			Random sample of 666 gold-standard 'Personal History' documents	Document-level Overall P=79% Overall	N/A

	social worker are assigned relation feature of ‘relation-other’. Full list of these occupations is available on request.				R=77%	
2	As above			Random sample of 200 ‘personal history’ documents	Document-level Overall P=77% Overall R=79%	N/A
3	As above	Random sample of 82 ‘personal history’ documents from records of patients aged >= 16 years – 40 document overall, 41 documents patient-only testing	Overall P=96% Patient-only occupation P=96%			
4	As above	Random sample of 116 documents (excluding ‘personal history’) from records of patients aged >= 16 years – 51 documents overall, 66 documents patient-only testing	Overall P=93% Patient-only occupation P=66%			

5	As above	Random sample of 166 documents annotated by app as 'other' occupation	Overall P=23%			

NOTES

This app performs best on personal history documents but can be used when text-mining from other free-text document types on CRIS.

Where the application can't identify the job title from the text, the feature is assigned as 'other'. Round 4 of testing showed that this annotation gives poor precision performance. This is because the application often assigns this feature to sentences which indicate work but are false positives (e.g. 'working on his anxiety'). It is advised that these annotations should be excluded from any analysis.

Production

- Run schedule – on request
- Version – 1

9. LIVES ALONE

Description

Application to identify instances of living alone.

Definition

Development approach: Rule-based***.***

The application identifies the following:

1. Lives on her own
   Who- none

2. She lives alone
   Who- She

3. He presently lives alone on 7th floor.
   Subject – He

4. His father lives alone.
   Subject – Father

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'lives on his/her own'), 'lives by him/herself', 'lives alone')

Search Terms (case insensitive)

Lives alone

Lives by himself

Lives by herself

Lives on his own

Lives on her own

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		Random sample of 100 - 1 presenting circumstances, 3 mental health	Overall P=97% Subject P=61%	Random sample of 100 – 50 attachments , 50 events	P=77% R=83%	lives on his/her own lives by him/herself

		formulations, 7 personal histories, 7 CAMHS events-clinical notes, 3 CCs correspondence - attached text, 1 mental health care plan, 32 correspondence - attached text, 46 events-clinical notes				lives alone

NOTES

Only three false positives in the annotated document, occurring as the mention of living alone was part of a list/was questioned and when a contradictory statement was used 'lives alone with….'. Subject precision was low because statements without an identifier e.g. he/she/ZZZ and just simply 'lived alone' were classified as 'none'. When these were excluded, precision rose to 83.5%. False positives in the non-annotated documents occurred due to uncertain references to living alone (similar to annotated) and certain negations. Positives not included (affecting recall) are mentions of the patient living 'independently'

Production

- Run schedule – on request
- Version - 1

10. DOMESTIC VIOLENCE

Description

Application to identify instances of domestic violence.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

The annotator is presented with a keyword, in a context, which they then should annotate as 'Positive' (relevant), 'Negative' (irrelevant) or 'Unknown' (at the machine learning stage, unknowns are counted as negative).

In the context of this application, annotators needed to annotate cases where the keyword 'DV' referred to any instance of *actual or alleged* domestic violence. This was conducted on the basis of a feminist empowerment model, where any and all allegations of domestic violence are taken seriously. Instances *do not* necessarily pertain to the patient, and can be historical or current.

Inclusion criteria involved any examples of violence in the context where the noun was contained "ZZZZZ was a victim of DV"; "X has a history of domestic violence"; "he experienced DV in the past".

Exclusion criteria were instances where the term "DV" or "domestic violence" did not describe any form of domestic violence (in most cases, this was where "DV" referred to "domiciliary visit"), for example "saw X on a DV yesterday". Furthermore, it was often not clear whether an allegation of domestic violence had taken place, for example: cases where domestic violence was denied "e.g. ZZZZZ denied that DV took place", "denies any domestic violence" were excluded.

Interrater reliability

Cohen's k = 87% (180 unannotated documents, search terms 'dv' and 'domestic violence')

Search Terms

domestic violence

DV

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 – CAMHS events-comments	P=94%	Random sample of 100 (50 events and 50	P=86% R=93%	'DV' 'domestic'

				corresponde nce)		

Production

- Run schedule – monthly
- Version – 2

11. LONELINESS

Description

Application to identify instances of loneliness.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive: Text indicates that the patient is lonely, or the patient confirms they have a sense/feeling of loneliness.
Negative: Patient is not lonely or denies being lonely.
Unknown: There is reference to loneliness but it does not relate to the patient themselves. Examples of this could be that the patient's family member is lonely; that they are participating in an activity on a ward to prevent boredom/loneliness. Instances where the EHR discusses the prevention of loneliness would be classified as unknown (however, if the text mentions "preventing further loneliness", this can be identified as positive, as it confirms that the patient has been lonely). Instances where a clinician suspects loneliness, or if there "might be loneliness/lonely" but it is not declared by, or agreed by the patient, would be classified as unknown. Forms: Loneliness if indicated on a form as a heading or question would be classified as unknown.

Interrater reliability

Cohen's k= 81% (100 unannotated documents, search terms 'lonely', 'loneliness')

Search Terms

lonely

loneliness

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100	P= 86%	Random sample of 100	P= 77% R= 98%	lonely loneliness
2	The application excludes: no lonel*, denied feeling lonely, not feel lonely,			Random sample of 100	P= 87% R= 100%	lonely loneliness

	mother is lonely, father is lonely, do you have long-term feelings of emptiness and loneliness					

Production

- Run schedule – monthly
- Version – 1

12. VIOLENCE

Description

Application to determine if mentions related to violence are being described as clearly relevant to the patient, and if they are also indicating that the patient is the *perpetrator*, or the *victim*; and if the type of violence relates to *physical*, *domestic*, or *sexual*. This is in essence 6 applications: violence status, domestic violence, sexual violence, physical violence, violence perpetration, violence victimisation.

Definition

Development approach: Rule-based***.***

Classification of past or present symptom: Both.

Classes produced:

1. affirmed/negated/irrelevant ("status"),
2. when affirmed: related to the patient being the perpetrator or victim (could be both)
3. when affirmed: domestic, sexual or physical (could also be more than one type)

Detailed annotation guidelines are available: https://docs.google.com/document/d/19OxdB0kMYsinPG-4xh-Ft0DZi-FJpgPQkBm4T_VkMXc/edit?usp=sharing

Interrater reliability:

Cohen's k

Status = 85% (2652 un-annotated documents)

Victimisation = 70% (2652 un-annotated documents)

Perpetration = 80% (2652 un-annotated documents)

Domestic = 74% (2652 un-annotated documents)

Physical = 60% (2652 un-annotated documents)

Sexual = 68% (2652 un-annotated documents)

Search Terms

% abus%

% assault%

% attack%

% beat%

% chok%

% fight%

% fought%

% hit%

% punch%

% push%

% rape%

% slap%

% strangl%

%strangul%

% struck%

% threw%

% violenc%

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		**Status** Random sample of 100 (25 attached texts, 2 CAMHS events, 2 CCS correspondence, 22 events-comments, 1 discharge plan, 2 mental state formulations, 1 nurse assessment notes, 2 risk event descriptions, 1 risk assessment tool descriptions, 9 risk assessment tool notes, 2 reason for referral, 2 single generic risk assessments, 1 mental health history, 1 physical health history, 14 summaries of need, 7 ward	P=94%	Random sample of 100 (50 events and 50 correspondence)	P=90% P=99%	abus* violenc* assault* attack* hit*

		progress notes, 5 ward rounds)				
2		**Status** Random sample of 50 patients	2019 P=89% Patient-level P=98% 2018 P=99% Patient-level P=98%			
3		**Physical** Random sample of 100 (37 attached texts, 3 CAMHS events, 2 CCS correspondence, 3 risk event description, 1 presenting circumstances, 1 mental statement formulation, 6 risk assessment tool risk factors, 1 social situation, 3 summary of needs, 13 summary of needs, 30 event comments)	P=86%	Random sample of 100 (50 events and 50 correspondence)	P=94% R=87%	abus* punch* assault attack* hit*
4		**Physical** Random sample of 50 patients	2019 P=95% Patient-level P=98% 2018 P=86% Patient-level P=98%			

5		**Sexual** Random sample of 100 (36 attachments, 3 CAMHS events, 1 mental health care plan, 1 discharge notification summary, 1 CCS correspondence, 8 ward progress notes, 2 ward rounds, 3 risk events, 3 risk assessment tool notes, 1 mental state formulation, 5 single generic assessments, 36 event comments)	P=84%	Random sample of 100 (50 events and 50 correspondence)	P=70% R=80%	abus* violenc* assault* attack* rape*
6		**Sexual** Random sample of 50 patients	2019 P=91% Patient-level P=100% 2018 P=95% Patient-level P=96%			
7		**Victimisation** Random sample of 100 (45 event-attachments, 3 CAMHS events, 1 CCS correspondence, 3 discharge notification summaries, 2 family histories, 1 SRA assessment note, 1 risk event	P=72%	Random sample of 100 (50 events and 50 correspondence)	P=78% R=93%	abus* violenc* assault* attack* hit*

		description, 6 risk assessment tool notes, 3 simple generic assessments, 3 ward progress notes, 32 event comments)				
8		**Victimisation** Random sample of 50 patients	2019 P=83% Patient-level P=82% 2018 P=76% Patient-level P=86% 2017 P=69% Patient-level P=70%			
9		**Perpetration** Random sample of 100 9; 24 event attachments, 2 discharge notification summaries, 1 CCS correspondence, 16 event comments, 2 mental state formulations, 1 POS Proforma, 1 Presenting circumstances, 3 risk events, 8 risk assessment tools, 5 single generic assessments, 18 summary of needs, 3 ward	P=69%	Random sample of 100 (50 events and 50 correspondence)	P=89% R=95%	abus* violenc* assault* hit* punch*

		rounds, 16 ward progress notes)				
10		**Perpetration** Random sample of 100 (28 attachment text, 2 CAMHS events, 2 CCS correspondences, 2 discharge notification summaries, 4 risk assessment tools, 1 single generic assessment, 2 summary of needs, 4 POSproforma, 2 wardround, 22 ward progress notes, 7 risk events, 1 personal history, 24 event comments)	P=71%	Random sample of 100 (50 events and 50 correspondence)	P=92% R=91%	abus* violenc* assault* hit* punch*
11		**Perpetration** Random sample of 50 patients	2019 P=76% Patient-level P=86% 2018 P=91% Patient-level P=72% 2017 P=90% Patient-level P=80%			
12		**Domestic** Random sample of 100 (43 attachment texts,	P=82%	Random sample of 100 (50 events and 50 correspondence)	P=89% R=95%	abus* violenc* assault* attack* hit*

		3 CAMHS events, 1 CCS correspondence, 4 discharge summaries, 1 discharge comment, 2 summary of needs, 1 risk event description, 1 personal history, 2 risk assessments, 5 ward progress notes, 37 event comments)				
13		**Domestic** Random sample of 50 patients	2019 P=92% Patient-level P=84% 2018 P=69% Patient-level P=96%			

Production

- Run schedule – on request
- Version – 1

INTERVENTIONS

1. COGNITIVE BEHAVIOURAL THERAPY (CBT)

Description

An application to identify instances of delivered sessions of Cognitive Behavioural Therapy (CBT).

Definition

Development approach: Rule-based***.***

Search Terms

1.1 Inclusions:

A session of CBT is defined as an **event (excluding ward progress notes)** having "CBT" or "Cognitive Behavioural Therapy" or "Cognitive Therapy" followed by "session", "assessment" or "follow up" plus the following variations specified below:

1.2 Assessment session:

Other terms that should be included

"CBT Assessment"	Assessment
"CBT: Ax"	Assessment
"Assessment and CBT in the same sentence"	Assessment
"Initial CBT appointment"	Assessment

1.2 Treatment session

Other terms that should be included:

"Attended for CBT"

"LICBT" & "session"

"CBT appointment"

"CBT appt"

"saw ZZZZZ for CBT"

"CBT: Seen"

"CBT: Reviewed"

"Session X of CBT"

"**X** CBT"

"**X**st CBT"

"CBT #**X**"

"CBT #**X**"

"S**X** CBT"

"session of CBT"

"continued with CBT"

"CBT psychology session"

"session X of CBT"

"Met with ZZZZZ to continue the CBT work."

"MIND WORKOUT (CBT GROUP)"

1.3 Follow up

"CBT follow up appointment" "CBT 12-month follow-up"

Alternative terms for CBT

"SX HICBT"

"SX LICBT"

Interrater reliability

N/A

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1				SQL CRIS Events where Event Type=Face-to-Face, group or phone, attended and formal therapy ID=CBT (derived table)	P=89%	CBT cognitive behavioural therapy
2				Raw table based on JAPE rules ran over CRIS events (GateDBCRIS.vw_gate_ cbt_session_session)	P=85% R=86%	CBT cognitive behavioural therapy
3				Raw table post-processed to exclude CBT session reference>200 characters from Event start (GateDBCRIS.vw_gate_ cbt_session_post_processed)	P=99% R=82%	CBT cognitive behavioural therapy
4				Post-processed and Structured Events combined (SQLCRISImprort.vw_ gate_ cbt_combined)	P=99%	CBT cognitive behavioural therapy
5				Materialised monthly version using the CBT combined view	P=99%	CBT

				(SQLCRIS_Common. dbo. tbl_cbt_combined_current)		cognitive behavioural therapy
6		Random sample of 100 – 9 CAMHS events-comments, 91 events - comments	P=89%			
7		Random sample of 100 – events - comments	P=57%			
8	Filter: NLP = 1 and start date >= 01-01-2015	Random sample of 100 – events - comments	P=98%			
9	Filter: NLP or event_rule = 1 and start date >= 01-01-2015	Random sample of 100 – events - comments	P=100%			

NOTES

Round 6

The main reason for the low precision is that the application description needs a direct label of 'cbt' or 'cognitive behavioural therapy'. However, most of the events-comments stated 'psychological session' or just mentioned 'session' with the intervention variable stating 'formal psychotherapy'. Precision would rise to 90% if we counted mentions of sessions and psychological assessment attendance as a CBT session. In some cases, the summary text stated 'CBT' while the event-comment did not mention CBT directly (just description of session). This was counted as positive although there were not many.

Round 7

One FP was due to the mention of not being a clear session and the other was where the mention was not an instance of the actual CBT session but a different session happening simultaneously with a family member.

Round 8

All instances were attended CBT sessions with 3 CBT assessments.

Production

- Run schedule - weekly
- Version -1

2. FAMILY INTERVENTION

Description

The application identifies instances of family intervention delivery.

Definition

Development approach: Rule-based***.***

The application will produce the following 6 features for each annotation: -

FI Session: Y/N

Session n: Session number

Stage: Assessment, first session, last, treatment, follow-up,

Subject: Both patient and carer/Carer/Patient but patient only relevant FI intervention for Behavioural Family Therapy (BFT). – Note if a single subject + patient then annotate as both ("ZZZZZ and carer") and if more than one other attendee then annotated as family ("ZZZZ, mum and sister").

Delivery: Individual Family/Multi Family – note Multi family groups are not generally practiced in the psychosis services but will be in the eating disorders service

Outcome: Attended, DNA, cancelled

Annotations for each group will be assigned independently of each other, e.g. in theory the same text could produce annotations from each group.

FI Session

Inclusions

A session of FI is defined as an event having "FI" or equivalent terms ("family intervention", "FI", "family therapy", "family work","family workshop","systemic work","systemic therapy", "family session", "FTWS", "Behavioural Family Therapy", "BFT", "BFI", "FIP") followed by "session" or equivalent terms ("appt", "Appointment", "Ass", "Assessment", "Reviewed", "Seen") and additional terms specified below.

Exclude "family meeting" and "carer" from NLP app but include in the heading section – exclude at the combined_view stage.

Note - FIP refers to Family Intervention in Psychosis

Assessment session

Other terms that should be included	Stage
"FI Assessment"	Assessment
"FI: Ax"	Assessment
"Assessment and FI in the same sentence"	Assessment

Treatment session

Other terms that should be included:

“Attended for FI”

“FI appointment”

“FI appt”

“saw ZZZZZ for FI”

“FI: Seen”

“FI: Reviewed”

“Session X of FI”

“X FI”

“Xst FI”

“FI #X”

“FI #X”

“SX FI”

“session of FI”

“continued with FI”

“session X of FI”

“Met with ZZZZZ to continue the FI work.”

Follow up

“FI follow up appointment”

“FI 12-month follow-up”

Exclusions

The following combinations below with FI in the same sentence are considered as exclusions. Note if the above inclusion criteria are met then this would be considered a positive hit independently of below but if only “next session” and FI were present in the same sentence this wouldn’t be annotated as a positive hit: -

“next session -/-” (day/month)

“next session 2nd”

“next session _._._” (day/month/year)

“Next session _._” (day/month)

“next appointment -/-” (day/month)

“next appointment 2nd”

“next appointment _._._” (day/month/year)

“Next appointment _._” (day/month)

“next appt -/-” (day/month)

“next appt 2nd”

“next appt _._._” (day/month/year)

“Next appt _._” (day/month)

Session n

Where a FI session has been indicated record the session number where specified. Note include first and last. Think about proximity – usually “Session x” but also examples of 1st session of FI, etc…

Other terms

“Final FI session”

“last FI Session”

“Final session of FI”

“last session of FI”

Stage

Assessment terms:

“FI Assessment”

“FI: Ax”

“Assessment” and “FI” in the same sentence

Some services e.g picup service has mid therapy assessment

Follow-up terms

“FI Follow up appointment”

“FI Follow up appt”

Subject

Inclusions

Both patient and carer

Carer/

Patient but patient only relevant for Behavioural Family Therapy (BFT) (only in psychosis services)

Delivery

Inclusions

Group or individual therapy

Outcome

Attended, DNA, cancelled by carer, cancelled by patient, cancelled by staff

Interrater reliability

Cohen's k = 88% (50 annotated documents - 25 events/25 attachments)

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 - event – clinical note	P=27%, Session P= 50% Session number P= 96%	Random sample of 100 - event – clinical note	P=77% R=87%	FI family intervention family therapy family work* systemic work systemic therapy family session FTWS behavioural family therapy BFI BFT FIP
2	Application excludes instances of '*meeting*' and '*carer*'	Random sample of 100 - 17 CAMHS events, 83 event-clinical notes	P=84%			

3	fi_term_exclude_for_testing=1	Random sample of 100 - 100 CAMHS events	P=93%			
4	fi_term_exclude_for_testing=1	Random sample of 100 - 100 CAMHS events	P=92%			
5	Filter: fi_term_exclude_for_testing=0, NLP=1 and structured_foram_therapy_FI_entry=0	Random sample of 100 - 100 CAMHS events	P=96%			
6	Filter: fi_term = Family Intervention	Random sample of 100	P=99%			

NOTES

Round 2

False positives occurred each time because the mention was not of an actual FI instance. They were comments on talking about referring to FI, or cancelled sessions. Also, mentions were discussions on what FI is without stating whether it was going to be undertaken by the patient/their family. Instances also included waiting for a referral or being on the waiting list without having undergone FI yet. Negatives also included discussion family meetings that were not therapeutic e.g. the logistics of the patients care plan. These also involved denying the idea of family therapy.

Post processing rules added on the most frequent false positives: not including ‘cancel’, ‘cancelled’, ‘DNA’ and ‘did not attend’. Recall was not tested with post processing rules and post processing precision was only measured on the annotated document.

Precision on non-annotated documents was much higher, as most of the positives related to actual FI instances rather than discussion of referral/assessment. Both documents were all event clinical notes.

Session number precision was high as only one event note gave the session number. The app produced ‘NULL’ as a response to each case, making it correct in all but that one mention.

Low session precision was mainly due to labelling sessions as ‘n’ rather than ‘y’. Due to unclear classification of positive instances, this is a hard outcome to determine. I measured this as ‘y’ being the actual note commenting on a therapy session, while ‘n’ was the patient/consultant briefly mentioning a previous session that would have been described more in detail in another clinical note.

Rounds 3, 4 and 5

Precision was good for both groups. Only 6 of those excluded (exclusion for testing=1 group) did not reference a ‘family meeting’. Therefore, these were consistently being excluded correctly. Instances where there were FPs were mentions of a family session, family work, family CBT session or ITP session. Precision for the included group (exclusion for testing=0 group) were consistent mentions of family work/family CBT session. False positives mainly related to home visits where FI was not specified, with one stating change to a family therapy appointment.

Code for post-processing

fi_term not like '%meeting%' and *fi_term* not like '%carer%'

Production

- Run schedule - weekly
- Version – 1

3. MEDICATION

Description

The Medication Application distinguishes between medications that are *currently* prescribed (i.e. at the time of the document was written) and medications prescribed to the patient *in the past*. The application ignores medications that might be prescribed in the future. This is because a clinician may write that a patient should be prescribed a certain drug if their condition worsens but that may never happen to the patient. The Medication application does not calculate daily dose of a drug, just the dose given at a single point in time.

The application output is linked to BNF codes to enable researchers to filter by drug class. *N.B.* Some drugs with antidepressant BNF codes appear more frequently as antipsychotics (e.g., flupentixol). Care should be taken when extracting patients who have ever used an antidepressant to ensure antipsychotic usage is not erroneously included. Corresponding dosage information is informative in determining whether a patient used a drug as an antipsychotic or as an antidepressant.

Definition

Development approach: Rule-based***.***

The application appears to preferentially detect medications:

(a) With corresponding dosage information

(b) Written in this format: 'Medication:' 'Current medications:'

Interrater reliability

N/A

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1				Random sample of 100 – 50 attachments, 50 events	P = Not sufficient positive instances to test **Recall** Current rx: overall (90%); attachments (94%); events (83%) Current rx or direct evidence of current use:	BNF antipsychotics

						overall (79%); attachments (90%); events (67%)	
2					Random sample of 100 – 50 attachmen ts, 50 events	P = Not sufficient positive instances to test **Recall** Current rx: overall (71%); attachments (53%); events (86%) Current rx or direct evidence of current use: overall (69%); attachments (53%); events (82%)	BNF antidepressants
3					Random sample of 100 – 50 attachmen ts, 50 events	P = Not sufficient positive instances to test **Recall** Current rx: overall (83%); attachments (89%); events (73%)	BNF antipsychotics
4					Random sample of 100 – 50 attachmen ts, 50 events	P = Not sufficient positive instances to test **Recall** Current rx: overall (71%); attachments (53%); events (86%)	

					Current rx or direct evidence of current use: overall (71%); attachments (53%); events (86%)	
5				Random sample of 50 (only if one mention per document)	**Precision** Drug=99% Dose=99% **Recall** Drug=88%	Amlodipine
6				Random sample of 200 – 100 attachments, 100 events	**PRECISION** **Attachments** Instance level - Ever used: 96%; Instance level - current use: 71%; document level - current Rx: 82% **Attachments filtered for present tense only** Instance level ever used - 96%; instance level current use 76%; document level current Rx 85% **Attachments with dosage** Instance level ever use - 97%; instance level current use - 76%; doc level current rx - 84%	Antipsychotics

					Events Instance level - ever used: 94%; instance level - current use: 85%; document level current Rx: 77% **Events filtered for present tense only** Same as without filtering **Events with dosage** Instance level ever use - 98%; instance level current use: 92%; doc level current rx: 87% **Dosage precision (including precision of unknowns)**: 94% **Tense precision**: 76% **RECALL** Not tested	
7				Random sample of 50 – 25 attachments, 25 events	**PRECISION** **Attachments** Instance level - ever used: 94%; current use: 84%; doc level - current use: 88%, current Rx: 88% **Attachments with dosage** Same as overall	Diabetic drugs with BNF code ‘060101*’ or ‘060102*’

					precision without dosage **Events** Instance level - ever used: 94%; current use: 82%; doc level - current use: 88%, current Rx: 73% **Events with dosage** Same as overall precision without dosage **Tense precision** Overall 73% (83% for present, 19% for past) **RECALL** Not tested	
8				Random sample of 20 – for patients with 1st prescription after 01.01.2007 for any of the medication terms	**PRECISION** **Antipsychotics** Document level - Ever use: 97% Document level - current use: 88% Patient level - Ever use: 99% Start date precision -Same day: 53% one week: 61% one month:63% three months: 66% **Antidepressants** Document level - ever use: 94% Document level - current use: 85%	Olanzapine Clozapine (filtered for dose info only) Risperidone Aripiprazole Quetiapine Sertraline Citalopram Mirtazapine Fluoxetine Venlafaxine Sodium valproate

					Patient level - ever use: 97% Start date precision: Same day:42% one week:43% one month:49% three months: 59% **Sodium Valproate** Document level - ever use: 90% Document level - current use: 80% Patient level - ever used: 99% Start date precision: Same day: 45% one week: 50% one month:50% three months: 50% **RECALL** Antipsychotics: Evidence of earlier use than the start date indicated by the app: 39% of records. App-detectable for 17% of these records. Antidepressants: 47% of records indicated an earlier start date 10% of these were app-detectable	

					Sodium valproate: 50% of records indicated an earlier start date 5% were app-detectable	

Production

- Run schedule – weekly
- Version - 2

4. SOCIAL CARE – CARE PACKAGE

Description

Application to identify instances of receiving current, recommended or planned general care package. This is a generic term relating to any social care intervention. This could be a specific type of social care (e.g. Meals on wheels, regular home care) or general mention of a package. 'Status' states whether patient currently has a care package, will get one in the future or that there is potential to receive it. 'Subject' states who the receiver of the care package is.

Definition

Development approach: Rule-based**.**

Classification of past or present instance: Both.

Classes produced**:** Positive, Negative and Unknown.

Positive mentions include: 'carer came twice that day as per the care package', 'should receive a package of care from next week', 'we have recommended the care package to be increased'.

Negative mentions include: 'refused the care package', 'tried to discuss having a package of care after discharge, but refused to converse'. Anything with suggested uncertainty of recommendation should be classed as negative.

Unknown mentions include: vague mentions that do not suggest that a care package is actually recommended.

Status definition

Current- is currently receiving a care package (present tense mentions).

Potential- recommended or discussions relating to possible care package. No exact plan to have it in the future but future discussions to take place.

Future- planned or arranged meals care package.

Subject definition

The individual who the care package relates to e.g. 'his wife', 'he', 'patient' or 'none' if the mention does not directly specify the individual.

Interrater reliability

Cohen's k = 95% (100 un-annotated documents, search term 'care package', 'package of care')

Search Terms

'care package'

'package of care'

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS

1		Random sample of 100 – (39 correspondence-attached text, 1 CCS correspondence-body text, 1 physical care plan, 1 discharge notification summary, 58 events-comments)	General P=91% General Status P=63% General Subject P=24% **Of those labelled general true positives (N=91)** Specific to Status P=68.1% Specific to Subject P=25.3%	Random sample of 100 – (50 events-comments, 50 correspondence-attached text)	P=88.1% R=72.8%	'care package' 'package of care'

NOTES

The few false positives raised were negation of 'no' care packages and mentions of care packages prior to admission (past tense: not identified by the app).

It was hard to identify a pattern amongst the false negatives, although the majority were potential/future mentions of 'arranging', or 'requiring' a 'comprehensive' care plan.

When precision of 'subject' and 'status' was re-tested on general true positives only, precision of subject went up to 25.3% and status precision to 68.1%, suggesting that applying the general rules to reduce general false positives could increase precision enough for status to possibly be used with extra modifications specific to that variable.

For 'status' specifically, most of the incorrect labels were when the output stated 'current' rather than 'future' mentions.

However, precision for 'subject' was low both tests due to the amount of 'none' output labels on documents. If this was automatically processed as 'patient', precision would increase significantly. Alternatively, this variable could be seen as unnecessary, as most documents would relate to the patient anyway.

Production

- Run schedule – on request
- Version - 1

5. SOCIAL CARE – HOME CARE

Description

Application to identify instances of home care/help. This is help by someone who comes to assist the patient with activities of daily living.

Definition

Development approach: *Rule-based.*

Classification of past or present instance: Both.

Classes produced: Positive, Negative and Unknown.

Positive mentions include: 'patient gets home help 3x daily', 'carers admit he can be agitated during the day at home', 'home care plan to start next week', 'meeting to discuss potential home care', 'suggested home help which the family are considering'.

Negative mentions include: forms, informal care not provided by a care service (informal carers, such as family members, friends or neighbours), ward round notes referring to care coordinator, 'home care FORMCHECBOX', 'carer support to be given to sister of ZZZ', 'refused home help', 'home help discussed but ZZZ does not want at this time', 'care coordinator contacted', 'ward round: care coordinator responded with…'.

Unknown mentions include: where it is unclear if current/potential home care is being given, this may be questions, mention of home treatment team without reference to home care or mention of past care without current/future care mentioned e.g: 'Does the patient have home care support?', 'home treatment team called', 'used to have home care support'.

Status definition

Current- is presently receiving home care/help.

Future- plans have been made for home care to occur.

Potential- home care has been suggested but there has been no planning of it being used for certain in the future.

Subject definition

The individual who the care package relates to e.g. 'his wife', 'he', 'patient' or 'none' if the mention does not directly specify the individual.

Time definition

This is how often the home care is received, outputs include day, week, month or NULL if not specified in the text.

Frequency definition

This refers to how often homecare occurs within the time frame (numeric output or NULL if not specified). For example, if homecare was received 3 times daily, the output would be: time= day, frequency=3.

Interrater reliability

Cohen's k = 95% (100 un-annotated documents, search 'home care', 'carer visits', 'carer support',

'home carer', 'home help', 'home helper', 'home carer had visited')

Search Terms

‘home care’

‘carer visits’

‘carer support’

‘home carer’

‘home help’

‘home helper’

‘home carer had visited’

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 – (42 correspondence-attached text, 2 CCS_correspondence - body text, 1 discharge summary, 1 physical care plan, 54 events-comments)	General P=80% General Status P=67% General Subject P=31% General Time P=68% General Frequency P=71% **Of those labelled general true positives (N=80)** Specific to Status P=83%	Random sample of 100 – (50 events-comments, 50 correspondence-attached text)	P=79% R=65%	‘home care’ ‘carer visits’ ‘carer support’ ‘home carer’ ‘home help’ ‘home helper’ ‘home carer had visited’

			Specific to Subject P=23% Specific to Time P=85% Specific to Frequency P=88%			

NOTES

Relating to general precision, the consistent patterns related to talking to a 'care coordinator' or 'care coordination' without explicit reference to home care. Also, reference to 'carer support' for the relative who is a carer. Informal carers are not labelled as positive in this app, as it is supposed to only reference a formal care service being used.

In terms of general recall, the only general pattern that could be seen was false negatives of 'home care' and 'home carer 'mentions.

Regarding specific precision of the other output variables, when precision was measured again on only general true positives for the specific output variables, precision rose. This suggests that putting in place general precision rules would allow the 'frequency', 'time' and 'status' variable output to be used.

Specifically for 'subject', if the output 'none' was automatically re-labelled 'patient', precision would rise significantly. However, as all documents relate to the patient anyway, perhaps this variable does not need to be used for the app.

Specifically for 'status,' the app usually incorrectly labelled 'potential' labels as 'current', although there were no similarities in the actual text string.

Production

- Run schedule – on request
- Version - 1

6. SOCIAL CARE – MEALS ON WHEELS

Description

Application to identify instances of receiving current or recommended meals on wheels (food delivery, usually from a private firm).

Definition

Development approach: Rule-based.

Classification of past or present instance: Both.

Classes produced: Positive, Negative and Unknown.

Positive mentions include: 'receiving MOW', 'recommended for MOW', 'planning to have meals on wheels from Monday', 'will arrange for ZZZ to have MOW', 'isn't eating his MOW', 'prefers the Wiltshire farms to the original MOW', 'discussed MOW', 'would be happy to have MOW'.

Negative mentions include: 'mow the lawn', spelling error 'he is mow in mood', 'refuses MOW', 'stopped having meals on wheels as he did not like the taste', 'used to have MOW, now stopped', 'FORMCHECKBOX' forms.

Unknown mentions include: 'MOW?', 'consider MOW', 'need to consider a plan that could partly include meals on wheels', 'will discuss MOW'.

Status definition

Current- is currently receiving MOW (present tense mentions). Current use relates to MOW delivery, if the individual refuses to eat/ignores the food, this would still be classed as a positive instance.

Potential- recommended, planned or arranged meals on wheels. Anything with suggested uncertainty of recommendation should be classed as negative.

Interrater reliability

Cohen's k = 95% (100 un-annotated documents, search 'MOW', 'meals on wheel*')

Search Terms

'MOW'

'meals on wheel*'

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		Random sample of 100 – (30 correspondence-attached text, 1 physical health care plan, 1 personal	General P=80% General Status P=66%	Random sample of 100 – (50 events-comments, 50 correspondence-attached text)	P=71% R=96%	'MOW' 'meals on wheel*'

		history, 68 events-comments)				

NOTES

General precision was reduced due to forms 'FORMCHECKBOX' and negations e.g. being offered but refusing/declining meals on wheels. Also, there were some past mentions prior to discharge.

The reduced precision in the non-annotated vs annotated documents may have been due to the search term 'MOW', as there were many false negatives of irrelevant mentions e.g. 'mow' instead of 'low' spelling error or 'mow the lawn' that was not present in annotated document testing. Similarly, like annotation testing, negations referring to refusing/declining MOW were also labelled as positives.

Regarding general recall, there were a few amount of false negatives, meaning no consistent pattern could be seen.

There were four other specific output variables; 'time', 'interval', 'name' and 'subject'. The 'subject' variable always relates to the patient therefore, I ignored this output. The other three variables only produced 'NULL' as an output, so this should also be ignored.

The reduced precision in current status was because it failed to identify potential mentions as potential, labelling them as current use (apart from in 2 instances). His could be referring to the 'set up', 'setting up', or it being 'discussed' or 'to start'.

Production

- Run schedule – on request
- Version - 1

OUTCOMES AND CLINICAL STATUS

1. BLOOD PRESSURE (BP)

Description

Application to identify instances of blood pressure scores in the format of overall score, systolic blood pressure score and diastolic blood pressure score.

Definition

Development approach: Rule-based**.**

Interrater reliability

Cohen's k = 98% (100 un-annotated documents - 50 events/50 attachments, search term 'bp')

Search terms (case insensitive)

blood pressure

bp

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1				Random sample of 100 events and attachments	**Precision** Overall: 98% Systolic: 98% Diastolic: 98% Full score: 98% Same day precision: 92% One week: 98% One month: 98% R=96%	blood pressure bp
2			Overall P=99% Systolic P=		Overall P = 98% Overall R = 85%	bp

			98% Diastolic P= 99%			

Production

- Run schedule – monthly
- Version - 1

2. BODY MASS INDEX (BMI)

Description

Application to identify body mass index (BMI) scores.

Definition

Development approach: Rule-based**.**

Interrater reliability

Search Terms (case insensitive)

bmi

Bmi

BmI

BMI

BMi

Body Mass Index

Body mass index

body mass index

Units for BMI: Kg/m2

Inclusions

Criteria	Examples
BMI and number in the same sentence	Bmi 45, bmi:46, Body Mass Index is 22.9, 16 BMI
BMI, number and units in the same sentence	Bmi 45 kg/m2, BMI 47 Kg/m2 , BMI 22.8 kg/m 2

Exclusions

Criteria	Examples
BMI and number in a sentence that indicates centile	Bmi centile 46, Bmi centile 77, He is on the 34th centile for BMI, BMI above 96th centile
BMI, number and units in the same sentence, bmi units are indicated wrong in the sentence	Her BMI is 48 kg, BMI: 22 kg, BMI/Weight : 103.2 kg
There is no score in the sentence, but there is an indication of the word BMI.	Record her weight to find out her BMI, BMI indicated that she was obese, Raised BMI, stable weight and BMI
BMI indicates as BMI range	BMI between 20.0 and 25.0, BMI within the healthy range 25.0 to 27.0

Features

BMI Score named as "BMI_Score" in the app has two features:

Kind (*examples in table below)*: >, < ,target, approx., +, over, assumed, aim, achieve, value of kind is blank if

Score: Numeric value of BMI

Values of Feature named as kind	**Example**
>	BMI greater than 17.5, BMI >17.5
<	BMI less than 18, BMI <19
target	Her target weight is 46 kg and BMI of 17, target BMI 17
approx	BMI of approx. 70
+	BMI 35+
over	BMI of over 28
assumed	Assumed BMI = 30.02
aim	Aiming for BMI 19
achieve	Agreed to achieve a BMI of 16
Otherwise value of kind is blank	BMI is 19

Examples

1) BMI is 24. 7 - Due to the space in between, app will only pick up score as 24 instead of 24.7

2) BMI is 48 kg - App will not pick this up.

3) BMI range between 24-25 - App will not consider this as a score

4) BMI is increasing - As there is no BMI score, app will not pick any score.

5) She is 40.66 kg and 153.5 height and is very skinny - As there is no mention of BMI score directly, app will not pick up any BMI score.

N.B. App will not pick up BMI if it is written in a table.

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1				Random sample of 100 – 50 attachments, 50 events	**Precision** Score precision: 89% (events: 89%; attachments 88%)	*bmi* *body mass index* *kg/m2*

					Date precision (automatic 22.47% penalty for FN) Same day: 66% (events: 70%; attachments: 63%) One week: 71% (events: 75%, attachments: 67%) One month: 72% (events: 78%; attachments: 67%) Three months: 73% (events: 78%; attachments: 69%) R = 78% (events: 80%; attachments 76%)	
2	App only accepts scores between 7 and 100 – see notes	Random sample of 100 – 40 attachments , 2 care plans mental health, 45 events, 1 mental state formulation, 2 risk events, 1 risk assessment tool CRIS risk plan, 1 risk assessment tool Risk	P=99%	Random sample of 100 – 50 attachments, 50 events	P=90% R=81%	bmi

		Factors, 2 summaries of need, 4 ward progress notes, 2 ward rounds				

NOTES

Code for post-processing

Bmi_score between 7 and 100

Production

- Run schedule - weekly
- Version - 1

3. BRAIN MRI REPORT VOLUMETRIC ASSESSMENTS FOR DEMENTIA

Description

Application for automated extraction of mentions of dementia-related volumetric assessments from plain text brain MRI reports

- This model has only been trained on brain MRI reports, where the clinical indication was for the investigation of dementia
- This had not been validated for use on other imaging reports, and for other clinical indications would be unlikely to return many results as the terms of interest are unlikely to be mentioned
- The terms of interest are unlikely to appear outside of the radiologists report, other than if they are copied into the patient notes and/or a clinical letter

Definition

Development approach: Machine-learning, based on spaCy python library

Classes produced: Patient referred for an MRI from memory clinic. Takes the plain text MRI report as input and identifies and returns labelled spans or token that are predicted as belonging to any of the below 6 classes:

- GVL = Global Volume Loss – Present
- NO_GVL = Global Volume Loss – Absent
- RVL = Regional Volume Loss – Present
- NO_RVL = Regional Volume Loss – Absent
- HVL = Hippocampal Volume Loss – Present
- NO_HVL = Hippocampal Volume Loss – Absent

For each span:

Text (string), label (String), score (float)

E.g., 'brain volume is normal' = NO_GVL
'Severe bilateral hippocampal atrophy' = HVL

The underlying span categorization approach allows multiple labels to be applied to the same span, e.g., 'no regional predominant or hippocampal atrophy' = NO_RVL and NO_HVL

Interrater reliability

Inter annotator agreement was done with one annotator as 'gold standard' and the other a assessed as if it was model output (as per token assessment in this case will be enormously skewed). Standard matching is exact token boundaries, relaxed matching is a labelled span having overlap (i.e., the presence of a span was agreed, but not the exact token boundaries which is important in this context as the labelled spans can be very long)

Standard matching:

Precision: 0.695, 95% confidence intervention 0.666, 0.722

Recall: 0.543, 95% CI 0.517, 0.569

F1 score: 0.610, 95% CI 0.591, 0.629

Relaxed matching:

Precision: 0.948, 95% CI 0.934, 0.960

Recall: 0.741, 95% CI 0.717, 0.763

F1 score: 0.832, 95% CI 0.817, 0.846

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		GVL F-score = 0.95	P=0.95		R= 0.95	
		NO_GVL F-score = 0.85	P=0.94		R=0.77	
		RVL F-score = 0.68	P=0.80		R=0.58	
		NO_RVL F-score = 0.92	P=0.91		R=0.93	
		HVL F-score = 0.89	P=0.90		R=0.88	
		NO_HVL F-score = 0.93	P=0.94		R=0.92	
		F1 score averaged over all 6 categories on holdout test set was 0.89	P=0.92		R=0.86	

Production

- Run schedule – yearly
- Version - 1

4. CHOLESTEROL

Description

To identify total cholesterol score of a patient, in clinical notes. Total cholesterol referred to as total cholesterol or serum cholesterol

Definition

Development approach: Rule-based***.***

Classes produced: In clinical notes, cholesterol level is referred as totals cholesterol score and serum cholesterol. Positive examples are:

Total cholesterol 06-Sep-2022 4.8 mmol/L

Serum cholesterol level 08-March-2001 3.8 mmol/L

Serum total cholesterol level 14-June-1998 6.3 mmol/L

Total cholesterol level 19-April-2010 5.3

19-March-2020 6 mmol/L serum cholesterol level

Negative examples are:
Serum cholesterol > 4.0 mmol/L

Interrater reliability

N/A

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 100 – 1 Addiction_Physical_Health_Assessment, 93 Attachment, 3 Event, 1 Triage_Assessment, and 2 Ward Progress Note	P=98%	Random sample of 100 – 50 attachments, 50 events	P=82% R=95%	total cholesterol serum cholesterol
		All patients, random sample of 100 – 94 Attachment, 1 Event and 5 Ward Progress Note	P=98%	Random sample of 92 – 84 attachments, 8 events	P=96% R=75%	total cholesterol

						serum cholesterol

Production

- Run schedule – monthly
- Version - 2

5. HBA1C

Description

The application will use a structured code to identify instances where HbA1c* and its results are found within CRIS from non-structured fields (i.e. case notes). This will help provide a clearer indication of how HbA1c is being recorded within CRIS.

*HbA1c can be obtained from a routine blood test and refers to glycated haemoglobin. It develops when haemoglobin, a protein within red blood cells that carries oxygen throughout your body joins with glucose in the blood, becoming 'glycated'. By measuring glycated haemoglobin (HbA1c), clinicians are able to get an overall picture of what our average blood sugar levels have been over a period of weeks/months. For people with diabetes, this is important as the higher the HbA1c, the greater the risk of developing diabetes-related complications. Therefore, it is important to ensure that this is being recorded and monitored effectively within SLaM as we know that those with psychosis are at a greater risk of diabetes.

Definition

Development approach: Rule-based***.***

Instances of HbA1c results were identified as follows:
Inclusion criteria:

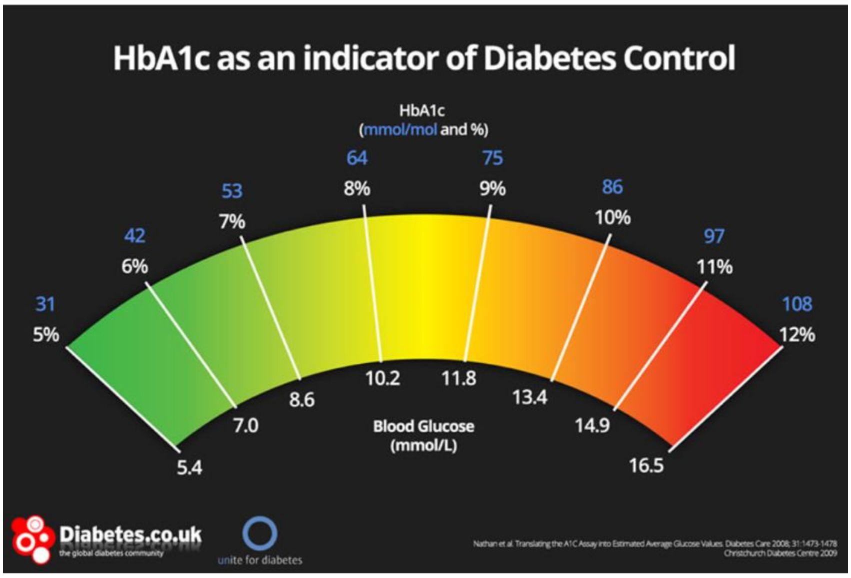
1) HbA1c score is present in format of e.g.
   a. HbA1c was 40, HbA1c 40, HbA1c was 40mmol/mol, HbA1c was 40mmol
   b. HbA1c was 15%
2) Decimals are allowed (e.g. 13.6)
3) HbA1c score relates to the patient only

Exclusion criteria:

1) HbA1c is mentioned without score e.g.
   a. HbA1c was measured and found to be within normal range
   b. HbA1c was measured on 11/11/19
   c. HbA1c 10/10/18

N.B: The application was not developed with upper or lower score limits. However, during testing anything lower than 3% or 9mmol and anything higher than 24% or 238mmol was excluded.

HbA1c	mmol/mol	%
Normal	Below 42 mmol/mol	Below 6.0%
Prediabetes	42 to 47 mmol/mol	6.0% to 6.4%
Diabetes	48 mmol/mol or over	6.5% or over



Interrater reliability

N/A

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 100 – 37 attachments, 6 discharge notification summaries, 20 events, 5 summaries of need, 27 ward progress notes, 5 ward rounds	P=94%	Random sample of 100 – 50 attachments, 50 events	P=92% R=76%	hba1c

Production

- Run schedule – monthly
- Version - 2

4. MINI-MENTAL STATE EXAMINATION (MMSE)

Description

This app identifies MMSE scores and returns information on:

- MMSE score (overall and subdivided into numerator and denominator)
- Associated date

Definition

Development approach: Machine-learning**.**

Numerator should be a number from 0 to 30 and denominator should always be 30. Date is identified in the format of DD/MM/YYYY.

Interrater reliability

Cohen's k = 90% (50 un-annotated documents - 25 events/25 attachments, search term '*MMSE*')

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		All patients, random sample of 100 (one document per patient)	Numerator P=97% Denominator P=98% Date P=68% - same day Date P=76% - one week Date P=81% - two weeks Date P=84% - one month			
2		Random sample of 100 - 2 mental formulation notes, 1 mental health care plan, 1	Overall P=95% Numerator P=99% Denominator P=99%	Random sample of 100 - 50 correspondence: attached text, 50	P=93% R=94%	*MMSE*

		discharge notification summary, 61 correspondence- attachments, 35 event comments	Date P= 86%	event comments		

NOTES

Overall, precision and recall were very good for each component. False positives were only seen when MMSE score had already been flagged in the document and it was raised twice, or another irrelevant score had been picked up. Occasionally, false positives occurred when the statement was questioning the MMSE score e.g. '/15, /20?'. Incorrect dates raised were often only a day off the correct date or occurred when there were multiple dates in the comments, and it was unclear what date defined what event.

Production

- Run schedule – weekly
- Version - 1

5. DIAGNOSIS

Description

Application to extract instances of diagnosis.

Definition

Development approach: Rule-based***.***

The main aim is to look for a standard or as close as possible to a definitive standard diagnosis:

1.) When reading through document, if you come across phrase(s) similar to the examples below:

……Diagnosis: Fxx.x diagnosis name……(this could be with or without the colon, or could even have several colons and/or other punctuation marks before they diagnosis name, following each

……Diagnosis Fxx.x diagnosis name……

……Diagnosis: diagnosis name……

……Diagnosis: Fxx.x……

Highlight this as 'Diagnosis' – please label the annotation just as I have specified it (i.e. with a CAPITAL D).

2.) The following features have been added under the Diagnosis annotation:

ICD10: if there is a name of a diagnosis, but no ICD10 code, find the ICD10 code and fill in under the feature ICD10

Diagname: if there is a diagnosis name then please copy this in the annotation feature. Please copy the exact diagnosis name even if it varies from the official ICD10 name.

Diffdiag – add this only if there is a differential diagnosis. This kind of diagnosis is often mentioned because usually most documents are trying to find out what the diagnosis is and in the process give a possible diagnosis which is vague or will not be the correct one eventually.

Nonpsychdiag – any definite diagnosis where the annotation does not come under the F group diagnosis. For example, COPD.

Interrater reliability

N/A

Search Terms (case insensitive)

Gazetteer of diagnoses and ICD10 codes.

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS

1				Random sample of 50 – 25 attachments , 25 events for each group	**Lifetime precision** F20/schizophrenia – 96% F20 – 100% SMI – 95% Schizoaffective – 80% Depression – 100% **Lifetime recall** F20/schizophrenia - 63% F20 - 65% SMI - 43% Schizoaffective - 29% Depression - 40%	F20* or schizophrenia F25 or schizoaffective or schizo-affective F32 or F33 or Depressi* schizophrenia, schizo-affective, bipolar, F20, F25, F33
2		All patients with primary diagnosis of learning disability in a structured field or unstructured text (*f7* or *learning dis*), random sample of 50	P=96%			
3	Refined exclusions	All patients with primary diagnosis of learning disability in a structured field or unstructured text (*f7*	P=93%			

		or *learning dis*), random sample of 100				

Production

- Run schedule - weekly
- Version - 1

6. TREATMENT-RESISTANT DEPRESSION

Description

Application to identify instances of treatment-resistant depression.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive and Unknown.

Positive annotations include has X year history of treatment resistant depression, problems with low mood (resistant depression), diagnosis: treatment resistant depression, resistant endogenous depression, suffers from chronic treatment resistant depression, referred for management of treatment resistant recurrent depression.

Unknown annotations include 'talked about ways in which they might resist allowing each other's depression to …', 'has a diagnosis of treatment resistant schizophrenia and depression', 'we discussed him enrolling for a study of treatment resistant depression', 'we talked about medication for treatment resistant depression', 'resisted antidepressant therapy for a number of years', 'needs an assessment to rule out treatment resistant depression', 'assess whether depression was resistant to mirtazapine', 'accepts that ECT is a strategy for treatment resistant depression'.

NB. There are no negative annotations i.e. no statements to say that someone did not have treatment resistant depression. On the database examined, the unknown annotations above were all labelled as 'negative', so this may need to be borne in mind when cross-evaluating.

Interrater reliability

Cohen's k = 85% (50 un-annotated documents - 25 events/25 attachments, search term 'resistant depression')

Search Terms (case insensitive)

depression [0-8 words in between] *resist*

resist [0-8 words in between] depression

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients with primary diagnosis code F32* or F33* in a structured field, random sample of 50	P=90%			

		(one document per patient)				
2		Random sample of 100 - 26 events, 39 attachments, 2 mental health care plan, 21 CCS correspondence, 12 ward progress	P=68%	Random sample of 100 – 50 attachments, 50 events	P=92% R=92%	resistant depression
3	Application excludes instances of '*i.e. treatment-resis*' (see notes)	Random sample of 100 - 31 events, 61 attachments, 2 mental health care plan, 3 CCS correspondence, 2 ward progress, 1 discharge notification	P=83%	Random sample of 100 – 50 attachments, 50 events	P=77% R=95%	resistant depression

NOTES

Precision is notably lower in the app output (annotated documents) (67%) compared to when the app is compared to 100 random documents (non-annotated documents) (92%). I suggest the reason for this being, the 100 'random' documents are extracted using the term %resistant depression%. The app's predefined search terms are: 'Depression [0-8 words] resist*' and 'Resist* [0-8 words] depression'. When these terms are used in conjunction with the extraction term '%resistant depression%' it is unsurprising that the precision is greater than the app using these search terms alone. 92% is therefore likely very optimistic and 67% is a more representative figure of the app's precision performance. 43% of the false positives raised by the app are due to this text string found at the bottom of the document: 'Criteria Checklist · Presenting problem is a moderate to severe mental health problem needing step 4 intervention, i.e. Treatment-resistant, recurrent or atypical depression'. Un-annotated documents precision has decreased, this may be due to the change in keyword from 'resistant depression' to '*resistant depression'. However, the majority of new false positives are due to the following expression: "Any other Asian backgroundInsufficient InformationAffective Disorders Unittreatment resistant depressionAffective Disorders". This expression is exclusively found in attachments. Prior to Post-processing rules added to application rules this expression was annotated as 'negative' but is now annotated as 'positive'. This could be resolved by excluding attachment documents containing the phrase '%Unittreatment%'. Other false positives included family history mentions, references to clinical trials investigating TRD and treatment resistant paranoid schizophrenia.

Code for post-processing

Contextstring not like ‘*i.e. treatment-resis*’

Production

- Run schedule – monthly
- Version - 1

7. BRADYKINESIA (DEMENTIA)

Description

To identify instances of bradykinesia in the context of dementia.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: positive, negative and unknown.

Positive annotations include presence of bradykinesia, motor symptoms – moderate bradykinesia L>R.

Negative annotations include absence of bradykinesia, he was moving easily in bed and transferring independently with no bradykinesia or tremor.

Unknown annotations include bradykinesia is a symptom of dementia, difficult to assess if it has caused any bradykinesia, SHO to look out for bradykinesia.

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'bradykine*')

Search Terms (Case insensitive)

bradykine

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 – 1 ward progress note, 13 correspondence-attached text, 86 events- clinical note	P=89%	Random sample of 100 - 50 events-clinical notes, 50 correspondence-attached text	P=91% R=84%	bradykine*

Production

- Run schedule – monthly
- Version - 2

8. TRAJECTORY

Description

Application to identify evidence from clinical text that something is getting better or getting worse.

Definition

Development approach: Rule-based (Jape rules)

Classification of past or present symptom: current

***Classes produced and search terms*:**

A1. The trajectory

Evidence from text that, at the time of writing, something has improved or deteriorated when compared with the past. Given the nature of clinical text the subject of change will normally relate in some way to the patient's health; either directly (e.g. "*Hallucinations have improved*") or indirectly (e.g. "*living conditions are worse than they were*') .

The app does ***not*** restrict output to these, i.e. references necessarily relating to the patient and their health. References to other types of change therefore may be included, e.g. "*Her piano playing has improved*'; "*His mother's eye-sight is better*".

Exclusion:

- Conditional change: e.g., "*His drinking is better when attends the session*"; "*If she smokes she feels worse*"
- Modals suggesting uncertainly or potential: e.g., "*she may have improved*"; "*his knee could deteriorate*"
- Past change, i.e. not likely to be recent or relating to 'now' (the time of writing): e.g., "*Symptoms improved when he was younger*"; "*The relationship with her mother deteriorated…*" ; "*He had been feeling worse…*"

Search terms and forms:

Improvement	Deterioration
To be better – incl. ***is*** better; ***has been*** better (recently)	To be worse
Other trigger verbs: -	
To improve – incl. improve***s;*** is improv***ing***; ***has*** improv***ed***; ***has been*** improv***ing***	To deteriorate
To get better	To get worse
To feel better	To feel worse
To sleep better	To sleep worse

To eat better	To eat worse
To drink better	To drink worse
To recover	To decline
	To worsen

A2. The subject

Subject 1: From a pre-defined list only (the subject gazetteer), the subject of change and its type / category is extracted

E.g. "*concentration has improved*"

Subject = 'concentration'; subject type = 'cognitive symptom'

The subject may be specific, e.g.

Subject 1	Subject type
self-care	ADL
appetite	Mental health symptom
etc.	

or non-specific, e.g.

Subject 1	Subject type
she	General
situation	General
etc.	

Subject 2: From the same list, a second subject is extracted if specified. E.g. "*stress and anxiety have deteriorated*".

Subject 1 = anxiety; subject 2 = stress

Subject feature(s) will be blank if:

- Subject is not directly in front of the 'trajectory' term(s), e.g. "*he's been working on his behaviour, which is getting better*"
- No subject is specified, e.g. "*Feels worse.*"
- The subject is not (yet*) in the subject gazetteer, e.g. "*His flat is better than the hostel he lived in previously*"

*new terms can be added to the subject gazetteer.

A3 Qualifiers

From a pre-defined list only, the qualifier feature identifies phrases that qualify the 'amount' of change, e.g. "*she feels a lot better*" – qualifier = 'a lot'; "*his eating has significantly improved*" – qualifier = 'significantly'; "*things aren't much better*" – qualifier = 'much'

A4 Negatives

There is a binary indication if the trajectory reference is negative, e.g. "*her memory hasn't got any worse*"

Performance

	Post-processing rules added to application	Annotated documents	Performance (annotated)	Un-annotated documents from keyword search in CRIS	Performance (un-annotated)	Keywords used for random extraction from CRIS
1		All patients, random sample of 100 (both Trajectory and Subject tested)	P=97% for trajectory (Trajectory is correct or not) P=98% for subject (When subject is not blank, subject of the trajectory is correct or not)		R=76% for Subject	

NOTES

Production

- Run schedule – on request
- Version - 1

9. TREMOR (DEMENTIA)

Description

Application to identify instances of tremor in patients with dementia.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Positive annotations include evidence of presence of tremor as a symptom or sign e.g. "There was evidence of a tremor when writing…", "…with a degree of resting tremor…"

Negative annotations include no evidence of tremor e.g. "There are no reports of any noticeable motor symptoms such as tremor…", "No dystonic movement or tremor".

Unknown annotations include "ZZZZ will be reviewed with regards to side effects and if there is no tremor then can have another 75mg of Paliperidone", "there is a family history of tremor".

Interrater reliability

Cohen's k = 100% (50 un-annotated documents - 25 events/25 attachments, search term 'tremor*')

Search Terms (case insensitive)

Tremor

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		All patients, random sample of 30 (one document per patient)	P=83%			
2		Random sample of 100 - 7 ward progress notes, 3 mental state formulations, 2 discharge summaries, 33 correspondence-attached text,	P=67%	Random sample of 100 – 50 attachments, 50 events	P=63% R=96%	tremor*

		55 events-clinical notes				
3		Random sample of 100 patients with dementia diagnosis - 11 ward progress notes, 2 mental state formulation, 47 attachments-attached text, 38 events-clinical notes, 1 css correspondence, 1 mental health care plan	P=88%	Random sample of 100 – 50 attachments, 50 events	P=83%, R=92%	tremor*

NOTES

False positives mainly occurred when tremors were distinctively not related to dementia. The main unrelated mention relating to anxiety, while others included medication, recreational drug or alcohol withdrawal or side effect. Negations were also labelled as positive, e.g. No tremors, no obvious tremor, denied getting tremors. False positives in the dementia-specific documents mainly occurred due to negations e.g. No tremor and denied any tremors. There were not enough false negatives to distinguish a pattern for recall, although this was high.

Production

- Run schedule – on request
- Version – 1

10. QT

Description

Application to identify instances of QT interval or Corrected QT interval, QTc.

Definition

Development approach: Machine-learning***.***

Classification of past or present symptom: Both.

Classes produced: Positive, Negative and Unknown.

Included:

- QTc now 497
- QTc 450,
- (Uncorrected QT 384ms)
- QTc – 404
- QTc: 421
- QTC interval of 430ms
- QTc=442ms

Excluded:

- increase the QTc by 1.3ms. *(Excluded since the patient's actual QTc not stated.)*
- QTc interval was less than 440ms. *(Excluded since the patient's actual QTc not stated.)*
- QT was given 7 days. *(Excluded since QT is referring to the patient.)*
- Date of birth: - QT- 1R "i '^n^^jSTtv^eN *(Excluded since a random mix of letters not clinically relevant.)*
- recommended QTC interval less than 440ms in men and less than 470ms in women. *(Excluded since it is stating the recommended ranges.)*

Interrater reliability

Search Terms (case insensitive)

QT

QTc

Followed by

=, <, >, -

Followed by

number

Performance

	Post-processing rules added to application	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted from keyword search in CRIS	Precision and recall (un-annotated)	Keywords used for extraction from CRIS
1		Random sample of 100 (24	P=94%		P=94% R=96%	qt*

		attachments, 1 CAMHS event, 4 discharge notification summaries, 23 events, 1 risk event, 1 single generic assessment, 1 summary of need, 29 ward progress notes, 6 ward rounds)				

Production

- Run schedule – on request
- Version – 1

MISCELLANEOUS

1. FORMS

Description

Application to identify documents that include form structures such as yes/no questions and other checklists.

Definition

Development approach: Rule-based***.***

Manual annotators look for forms as stand-alone documents or as part of the wider document text and annotate either positive (that text corresponds to a form and therefore the document contains a form or is a form), or negative (that text doesn't correspond to a form and therefore the document neither contains a form nor is a form).
A form is identified based on:
1) Presence of the term 'form' within document heading
2) Presence of yes/no questions
3) Presence of checkboxes

Forms are identified as such even if they are not filled in, are part of a letter or email or correspond to symptom checklists.

The following rules were applied by the app to determine the presence of a form.
-Parse the HTML Text, identifying the text within tags.
-Identify the term "Form" within heading, paragraph of emboldened tags with lengths of less than 80
- Identify the presence of yes/no questions
-Identify the presence of check boxes
-Identify unique text to particular forms (this is an evolving part of the app that is updated as more information becomes available).
-Identify terms to exclude the document from being a form. These are terms that indicate the document may be a letter or an email (such as a greeting to open or close a letter, or terminology to indicate an email reply, such as "re:".

Interrater reliability

IRR = 92% - Cohen's k could not be computed

Performance

	Post-processing rules added to application	**Annotated documents identified by the application**	**Precision and recall (annotated)**	**Un-annotated documents extracted from keyword search in CRIS**	**Precision and recall (un-annotated)**	**Keywords used for extraction from CRIS**
1		Random sample of 100 attachments – attachment text	P=100%	Random sample of 100 events and attachments	P=100% R=76%	*formcheckbox* *form 435* *mri request form*

Production

- Run schedule – on request
- Version – 1

2. QUOTED SPEECH

Description

Application to extract text within quotation marks.

Definition

Development approach: Rule-based***.***

Rules:

Regular expression matching is used to identify text occurring within matching quotation mark pairs [(" & "), (' & '), (“ & ”), (‘ & ’)] in the EHR. To avoid mistaking apostrophes used in contractions for the start of quoted phrase, a quote followed by a sequence (‘c’, ‘d’, ‘e’, ‘m’, ‘n’, ‘s’, ‘t’ , ‘ve’, ‘re’, ‘s’, ‘ll’, ‘all’) was treated as an apostrophe not a quote. A similar check was performed for end quotes. Once a quoted phrase was identified, any sub-quotations occurring within that quote were assumed to be part of the larger quotation. The length of quoted phrases was allowed to vary from one word to more than a paragraph; however, a maximum length of 1,500 characters was applied to avoid extracting the entire text where a quote was not properly closed. Phrases that consisted only of emails or starting with “https” were removed using standard regular expression pattern matching and substitution procedures.

Annotation guidelines:

Manual annotation involves identifying the full scope of human identifiable quoted speech.
Manual annotations that are deemed to be quoted speech if they include:

- Matched pairs of quotation marks as specified by the quotation algorithm [(" & "), (' & '), (“ & ”), (‘ & ’)].
- Mismatched quotation mark pairs e.g. “hello’
- Quotation mark pairs that start with the end quote e.g. ”goodbye“
- Incorrect/unusual quotation mark types that are not specified in the algorithm e.g. backticks in `x`
- Identifiable quotations that have no end quotation mark e.g. “I am always sad. The patient’s general mood was…

Manual annotations that are **not** deemed to be quoted speech if they include:

- Emails or URLs (https)

Interrater reliability

N/A

Performance

	Post-processing rules added	Annotated documents identified by the application	Precision and recall (annotated)	Un-annotated documents extracted	Precision and recall (un-annotated)	Keywords used for extraction from CRIS

	to application			**from keyword search in CRIS**		
1		Random sample of 100 - 3 CCS correspondence -attached text, 29 correspondence -attached text, 68 events-clinical note	P=82%			

Production

- Run schedule – on request
- Version – 1