



King's Research Portal

DOI:
[10.3791/60794](https://doi.org/10.3791/60794)

Document Version
Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Wang, T., Oliver, D., Msosa, Y., Colling, C., Spada, G., Roguski, Ł., Folarin, A., Stewart, R., Roberts, A., Dobson, R. J. B., & Fusar-Poli, P. (2020). Implementation of a real-time psychosis risk detection and alerting system based on electronic health records using cogstack. *Journal of Visualized Experiments*, 2020(159), [e60794]. <https://doi.org/10.3791/60794>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

1 **TITLE:**

2 Implementation of a Real-Time Psychosis Risk Detection and Alerting System Based on Electronic
3 Health Records using CogStack

4
5 **AUTHORS AND AFFILIATIONS:**

6 Tao Wang¹, Dominic Oliver², Yamiko Msosa¹, Craig Colling³, Giulia Spada², Łukasz Roguski⁴, Amos
7 Folarin¹, Robert Stewart^{3,5}, Angus Roberts^{1,3}, Richard J.B. Dobson^{1,3,4,6}, Paolo Fusar-Poli^{2,3,7,8}

8
9 ¹Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and
10 Neuroscience, King's College London, London, UK

11 ²Early Psychosis: Interventions and Clinical-detection (EPIC) lab, Department of Psychosis Studies,
12 Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

13 ³National Institute for Health Research, Maudsley Biomedical Research Centre, South London and
14 Maudsley National Health Service (NHS) Foundation Trust, London, UK

15 ⁴Institute of Health Informatics, University College London, London, UK

16 ⁵Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience,
17 King's College London, London, UK

18 ⁶Health Data Research UK London, University College London, London, UK

19 ⁷OASIS service, South London and Maudsley NHS Foundation Trust, London, UK

20 ⁸Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy

21
22 tao.wang@kcl.ac.uk

23 dominic.a.oliver@kcl.ac.uk

24 yamiko.msosa@kcl.ac.uk

25 Craig.Colling@slam.nhs.uk

26 giulia.spada@kcl.ac.uk

27 l.roguski@ucl.ac.uk

28 amos.folarin@kcl.ac.uk

29 robert.stewart@kcl.ac.uk

30 angus.roberts@kcl.ac.uk

31 richard.j.dobson@kcl.ac.uk

32 paolo.fusar-poli@kcl.ac.uk

33
34 **KEYWORDS:**

35 Psychosis, psychosis risk calculation, electronic health records, real-time alerting, health
36 informatics, clinical decision support, prevention

37
38 **SUMMARY:**

39 We demonstrate how to deploy a real-time psychosis risk calculation and alerting system based
40 on CogStack, an information retrieval and extraction platform for electronic health records.

41
42 **ABSTRACT:**

43 Recent studies have shown that an automated, lifespan-inclusive, transdiagnostic, and clinically
44 based, individualized risk calculator provides a powerful system for supporting the early

45 detection of individuals at-risk of psychosis at a large scale, by leveraging electronic health
46 records (EHRs). This risk calculator has been externally validated twice and is undergoing
47 feasibility testing for clinical implementation. Integration of this risk calculator in clinical routine
48 should be facilitated by prospective feasibility studies, which are required to address pragmatic
49 challenges, such as missing data, and the usability of this risk calculator in a real-world and
50 routine clinical setting. Here, we present an approach for a prospective implementation of a real-
51 time psychosis risk detection and alerting service in a real-world EHR system. This method
52 leverages the CogStack platform, which is an open-source, lightweight, and distributed
53 information retrieval and text extraction system. The CogStack platform incorporates a set of
54 services that allow for full-text search of clinical data, lifespan-inclusive, real-time calculation of
55 psychosis risk, early risk-alerting to clinicians, and the visual monitoring of patients over time.
56 Our method includes: 1) ingestion and synchronization of data from multiple sources into the
57 CogStack platform, 2) implementation of a risk calculator, whose algorithm was previously
58 developed and validated, for timely computation of a patient's risk of psychosis, 3) creation of
59 interactive visualizations and dashboards to monitor patients' health status over time, and 4)
60 building automated alerting systems to ensure that clinicians are notified of patients at-risk, so
61 that appropriate actions can be pursued. This is the first ever study that has developed and
62 implemented a similar detection and alerting system in clinical routine for early detection of
63 psychosis.

64

65 **INTRODUCTION:**

66 Psychotic disorders are serious mental health illnesses that lead to difficulties in distinguishing
67 between the internal experience of the mind and the external reality of the environment¹, as well
68 as a higher than average risk of self-harm and suicide². Under standard care, these disorders
69 result in major public health impact with a significant health and economic burden on individuals,
70 families and societies worldwide³. Early interventions in psychosis can improve outcomes of this
71 mental disorder⁴. In particular, detection, prognostic assessment and preventive treatment of
72 individuals who are at clinical high risk for developing psychosis (CHR-P)⁵ provides a unique
73 potential to alter the course of the disorder, thereby improving the quality of life for many people
74 and their families^{3,6,3}. CHR-P individuals are help-seeking young people presenting with
75 attenuated symptoms and functional impairment⁷: their risk of developing psychosis is 20% at 2-
76 years⁸ but it is higher in some specific subgroups^{9,10}. Despite some substantial advancements,
77 the impact of preventive approaches in routine clinical practice is limited by the ability to detect
78 most individuals who are at-risk¹¹. Current detection methods are based on help-seeking
79 behaviors and referrals on suspicion of psychosis risk; these methods are highly inefficient in
80 handling a large number of samples¹¹. Thus, the scalability of current detection methods to the
81 vast majority of the at-risk population is quite limited¹². In fact, only 5% (standalone specialized
82 early detection services) to 12% (youth mental health services) of individuals at-risk of developing
83 a first psychotic disorder can be detected at the time of their at-risk stage by the current
84 detection strategies⁶.

85

86 To extend the clinical benefits of the preventive approaches in a larger number of at-risk
87 individuals, we developed an automated, lifespan-inclusive (i.e., across all ages), transdiagnostic
88 (i.e., across different diagnoses)¹³, clinically-based individualized risk calculator, which can detect

89 individuals at-risk of psychosis in secondary mental health care at scale, beyond those meeting
90 CHR-P criteria¹⁴. This risk calculator used a Cox proportional hazard model to predict the risk of
91 developing a psychotic disorder over six years from five routinely collected clinical variables
92 selected a priori, in line with methodological guidelines¹⁵: age, gender, ethnicity, age-by-gender
93 and primary index diagnosis. These clinical variables were selected based on a priori knowledge
94 obtained from meta analyses^{16,17}, as recommended by the state-of-the-art methodological
95 guidelines¹⁵. The number of predictors is limited to preserve the Event Per Variable ratio and
96 minimize overfitting biases; including too many variables without a priori filter leads to overfitting
97 problems and poor prognostic accuracy¹⁸. The method used to develop this model provides
98 similar prognostic accuracy to automatic machine learning methods¹⁸. Parameters of the Cox
99 model were estimated based on a retrospective de-identified cohort from the South London and
100 Maudsley National Health Service Foundation Trust (SLaM)¹⁹. SLaM is a National Health Service
101 (NHS) mental health trust that provides secondary mental health care to a population of 1.36
102 million individuals in South London (Lambeth, Southwark, Lewisham and Croydon boroughs), and
103 has one of the highest recorded rates of psychosis in the world²⁰. All data used in the model
104 development were extracted from the Clinical Record Interactive Search (CRIS) platform, a digital
105 case register system, which provides researchers with retrospective access and analysis of
106 anonymized clinical records¹⁹. The clinical information in CRIS is extracted from a bespoke
107 Electronic Health Record (EHR) system, at SLaM, called electronic Patient Journey System (ePJS).
108 SLaM is paper-free and ePJS represents the standard data collection platform for clinical routine.
109 Thus, the transdiagnostic risk calculator leverages EHRs and has the potential to automatically
110 screen large EHRs of patients accessing secondary mental healthcare, to detect those who may
111 be at-risk of psychosis. The algorithm of this transdiagnostic risk calculator has been published
112 previously^{6,14,21}. The transdiagnostic risk calculator has been externally validated in two NHS
113 Foundation Trusts^{14,21} and optimized²², demonstrating its adequate prognostic performance and
114 generalizability across different populations.

115

116 According to methodological guidelines on the development of a risk prediction model^{15,23}, the
117 next step after model development and validation is to implement the prediction model in
118 routine clinical practice. Implementation studies are usually preceded by pilot or feasibility
119 studies that address potential pragmatic limitations associated with the use of risk algorithms in
120 clinical practice. For example, required data for running a calculator, such as age, gender and
121 ethnicity, may not be available at the date of diagnosis or updated later. Effective methods for
122 handling missing data and synchronizing frequent updates in real-time data streams should be
123 considered to obtain the most reliable prediction results in an implementation. Furthermore,
124 since the initial development of the risk calculator was based on retrospective cohort data, it is
125 not known whether it can be used in a real-time data stream that is typical of a real-world clinical
126 setting. Another challenge is ensuring that relevant clinicians receive the recommendations
127 generated by the risk calculator within an appropriate time frame and within a shared and
128 accepted communication pathway.

129

130 To overcome these limitations, we have completed a feasibility implementation study employing
131 the individualized transdiagnostic risk calculator. The study included two phases: an in vitro phase
132 that was conducted using data from the local EHR, without contacting clinicians or patients, and

133 an in vivo phase, which involved direct contact with clinicians. The in vitro phase had two
134 manifold aims: (i) to address implementation barriers according to the Consolidated Framework
135 for Implementation Research (CFIR)²⁷ and (ii) to integrate the transdiagnostic risk calculator into
136 the local EHR. Implementation barriers included the communication of risk outcomes to
137 clinicians. In SLaM, all patients are invited to register for Consent for Contact (C4C), which
138 indicates their willingness to be contacted for research, without affecting the quality of care. This
139 reduces the ethical issues relating to contacting patients. Further to this, working groups with
140 clinicians aided tailoring of how this information was communicated. During the in vivo phase
141 (May 14th 2018 to April 29th 2019), all individuals (i) older than 14 years (ii) who were accessing
142 any SLaM service (boroughs of Lambeth, Southwark, Lewisham, Croydon), (iii) receiving a first
143 ICD-10 index primary diagnosis of any non-organic, non-psychotic mental disorder (with the
144 exception of Acute and Transient Psychotic Disorders; ATPD), or a CHR-P designation and (iv) with
145 existing contact details were deemed eligible. During the in vivo phase, new patients accessing
146 SLaM each week were automatically screened for their psychosis risk, and those with having a
147 risk greater than a certain threshold were detected. The research team then contacted the
148 patients' responsible clinicians to discuss further recommendations and eventually suggest a
149 further face to face assessment⁶. If those assessed were considered to meet CHR-P criteria, they
150 were referred to specialist CHR-P services, such as Outreach and Support in South London
151 (OASIS)²⁸. This would result in improved detection of individuals prior to the onset of a psychotic
152 disorder and provide a significant opportunity for altering the course of the disorder. Crucially,
153 this feasibility study involved the full integration of the calculator into the local EHR system, which
154 is the topic of the current article. The full protocol of this feasibility study, including an overview
155 of the plan for evaluating the proposed research, details on managing data security and ethical
156 issues, has been presented in our previous work⁶. The current article, as a part of the feasibility
157 study⁶, selectively focuses on presenting the technical implementation of a real-time psychosis
158 risk detection and alerting system based on the local EHR data. More specifically, the aim of this
159 study is to investigate the technical feasibility of this risk calculator in timely detecting at-risk
160 patients as soon as they access a secondary mental healthcare service. The full results of the
161 feasibility study, in terms of clinicians' adherence to the recommendations made by the risk
162 calculator, will be presented separately. A comprehensive evaluation of the effectiveness of the
163 proposed research, which requires randomized designs, is outside the scope of the current
164 research program. To our best knowledge, this is the first method describing the implementation
165 of a risk calculator based on live EHR data for early detection of psychosis.

166
167 Our approach to psychosis risk detection and alerting takes advantage of the CogStack platform.
168 The CogStack platform is a lightweight, distributed, and fault-tolerant information retrieval and
169 text-extraction platform²⁴. This platform consists of three key components: 1) the CogStack
170 Pipeline that uses the Java Spring Batch framework to ingest and synchronize data from a pre-
171 defined data source (both structured and unstructured EHR data in multiple formats such as
172 Word, PDF files and images) to a predefined data sink in real time; 2) Elasticsearch, a search
173 engine allowing for storage and querying of the full text of EHR data, as well as providing various
174 application programming interfaces (APIs) to embed advanced analytics into the engine; and 3)
175 Kibana, an interactive, web-based user interface that allows users to query data in Elasticsearch,
176 build visualization dashboards and set alerting on anomalies or other patterns of interest from

177 data. Moreover, CogStack incorporates the ability to alert clinicians to potential problems by
178 Email and SMS (text), allowing clinicians to receive timely notifications about at-risk patients
179 reported by the risk calculator.

180

181 We present a model of psychosis risk detection and alerting based on ePJS at SLaM, leveraging
182 the CogStack platform. Compared with the CRIS platform that provides a mechanism for
183 retrospective access to de-identified health records from ePJS on a weekly basis¹⁹, the CogStack
184 platform at SLaM enables access to an identifiable EHR in real time, bringing the alerting closer
185 to the point-of-care and the risk prediction in a prospective design, although both the CRIS and
186 CogStack platforms use data sourced from ePJS in SLaM. In the section that follows, we provide
187 details of the key steps in our approach, including preparing source data from the EHR, ingesting
188 the source data into the CogStack platform to enable full-text search via Elasticsearch, running
189 the psychosis risk calculator using a Python daemon thread, and setting interactive visualizations
190 and real-time risk alerting via the Kibana user interface. Any researcher who aims to build a real-
191 time risk detection and alerting system based on EHR data can follow the approach and its
192 reference implementation. As we shall elaborate below, the proposed method exploits open-
193 source, lightweight techniques with high flexibility and portability. This enables the risk calculator
194 to be run in various locations and shows a high applicability to other risk estimation algorithms.
195 Moreover, the method works as a straightforward approach to enhance the risk detection and
196 alerting functionalities of an EHR embedded in a general healthcare system.

197

198 **PROTOCOL:**

199

200 This study was approved by East of England - Cambridgeshire and Hertfordshire Research Ethics
201 Committee (Reference number: 18/EE/0066).

202

203 NOTE: We have developed this protocol based on the CogStack platform and the Python
204 programming language. This system requires Docker (more specifically, Docker Compose
205 <https://docs.docker.com/compose/>), Anaconda Python
206 (<https://www.anaconda.com/distribution/>) and Git (<https://git-scm.com/downloads>) pre-
207 installed on a device. The commands provided in this protocol are based on the Linux
208 environment. In the following, we provide the details of preparing source data from an EHR
209 database, ingesting the data to CogStack platform, and setting up a real-time risk calculation and
210 alerting system for psychosis based on the CogStack platform. Moreover, an online version of the
211 risk calculator was developed to facilitate numeric calculation of the probability of an individual
212 developing psychosis in secondary mental health care on <http://www.psychosis-risk.net>.

213

214 **1. Source data preparation**

215

216 NOTE: In most use cases, CogStack ingests source data from a specified database view that can
217 combine data from one or more source database tables, where a view is a searchable object in a
218 database that contains the result set of a stored query on the data. The setup of the ingesting
219 view is tailored by the specific use cases and deployment settings of a health record database
220 system. This protocol is developed based on a psychosis risk calculator developed and externally

221 validated twice by Fusar-Poli et al.^{14,21} and as part of a pilot implementation feasibility study⁶.
222 The protocol is based on an EHR database deployed with Microsoft SQL Server 2014.

223

224 1.1 Create a view object (called "*vwPsychosisBase*" in this protocol) in an existing EHR database
225 system to join necessary information of patients for psychosis risk calculation and alerting. Make
226 sure that this view includes all patients receiving a first primary diagnosis of non-organic and non-
227 psychotic mental disorder (recorded by the International Statistical Classification of Diseases and
228 Related Health Problems, Tenth Revision [ICD-10]), as defined in the original model^{14,21}.

229

230 1.2 Ensure that each record in the view involves three types of patient information: 1) the first
231 primary diagnosis of a patient in the EHR system, including ICD-10 diagnosis index (diagnoses
232 were clustered together into the following ten clusters: acute and transient psychotic disorders,
233 anxiety disorders, bipolar mood disorders, childhood and adolescent onset disorders,
234 developmental disorders, nonbipolar mood disorders, mental retardation, personality disorders,
235 physiological syndromes, substance use disorders) and diagnosis date; 2) a patient's demographic
236 data, including gender, ethnicity and date of birth; and 3) the most recent contact information of
237 care team for a patient, such as details of general practice (GP), consultants and care
238 coordinators. The first two types of information are vital for the psychosis risk calculator^{14,21}, and
239 the third type of information is to enable timely risk alerting.

240

241 1.3 Make sure that each record in the view has a unique identifier (e.g., "*patient_id*" used in this
242 protocol).

243

244 1.4 Select the last update timestamps of all source information related to a record in the view
245 (e.g., the last update times of a patient's demographic information and the patient's first primary
246 diagnosis information), and choose the latest timestamp as the last update date and time for the
247 record in the view (denoted as "*etl_updated_dttm*" in this protocol). The last update date and
248 time of a record allows CogStack to synchronize updates in the database, such as new and
249 updated records.

250

251 **2. Data ingestion**

252

253 2.1 Download or clone the code repository from Github ([https://github.com/cogstack-](https://github.com/cogstack-slam/psychosis)
254 [slam/psychosis](https://github.com/cogstack-slam/psychosis)) or by typing "`git clone https://github.com/cogstack-slam/psychosis.git`" in a
255 terminal window. The downloaded folder contains the code for psychosis risk calculation and
256 configuration files for deploying a CogStack instance.

257

258 2.2 Go to the "*cogstack_deploy/cogstack/*" directory and modify "*psychosis.properties*" to
259 configure CogStack Pipeline for data ingestion. Modify the settings of section "*SOURCE: DB*
260 *CONFIGURATIONS*" based on the EHR database setup, including specifying the IP address of the
261 database server, database name, database username and password. Modify the view name (i.e.
262 "*vwPsychosisBase*") and field names (e.g., "*patient_id*" and "*etl_updated_dttm*") if necessary. In
263 case of error in configuring this file, follow the instructions at
264 <https://cogstack.atlassian.net/wiki/spaces/COGDOC/pages/38043684/Quickstart>.

265
266 2.3 Go to the “*cogstack_deploy/common/elasticsearch/config/*” directory and modify the section
267 “*xpack.notification.email.account*” in the “*elasticsearch.yml*” file to configure an Email address
268 for sending alerts. A detailed instruction for Email configuration can be found on
269 <https://www.elastic.co/guide/en/kibana/6.4/watcher-create-threshold-alert.html>.

270
271 2.4 Go to the “*cogstack_deploy/*” directory and type “*docker-compose up*” to run the CogStack
272 platform. Execute this command with root access. If the process is completed successfully, there
273 will be printed status logs of the currently running services, including CogStack Pipeline,
274 Elasticsearch and Kibana, in the terminal. As a result, all data and updates in the source database
275 view will be timely ingested to an Elasticsearch index called “*psychosis_base*” in the CogStack
276 platform.

277
278 2.5 Open a web browser and access Kibana user interface by typing “<http://localhost:5601/>” (or
279 replacing “*localhost*” with a specific IP address of the server running the CogStack platform). For
280 the first time accessing Kibana, click the **Management** tab and **Index Patterns** tab to specify an
281 Elasticsearch index that one wants to access with Kinaba. Type “*psychosis_base*” in the “*Index*
282 *pattern*” field and click **Next step**. Select “*etl_updated_dttm*” for the “*Time Filter*” field name and
283 click **Create index pattern** to add the “*psychosis_base*” index pattern for Kinana.

284
285 2.6 Once Kibana is connected to the Elasticsearch index (i.e., “*psychosis_base*”), search and
286 browse the source data interactively through the “*Discover*” page. Kibana allows non-technical
287 users to search for both structured metadata and free text. Detailed instructions of using
288 “*Discover*” are available on <https://www.elastic.co/guide/en/kibana/6.4/discover.html>.

289 290 **3. Risk calculation**

291
292 3.1 Open a new terminal window and go to the “*psychosis/*” directory. Install all required Python
293 packages (including “*elasticsearch*”, “*elasticsearch_dsl*”, “*pandas*” and “*numpy*”) used in the risk
294 calculator by typing “*conda install package-name*” or “*pip install package-name*” in the terminal.

295
296 3.2 Type “*python risk_calculator.py*” to run the psychosis risk calculator. If the process is
297 completed successfully, logs of the risk calculation will be printed in the terminal and the risk
298 results will be stored in a new Elasticsearch index called “*psychosis_risk*” within the CogStack
299 platform.

300
301 3.3 Check the risk results by using the Kibana interface. Similar to Steps 2.5 and 2.6, add a new
302 index pattern “*psychosis_risk*” to connect Kinbana with the “*psychosis_risk*” index, and explore
303 the risk results through the “*Discover*” page. To facilitate identifying new patients at-risk, use
304 “*first_primary_diagnosis_date*” as the “*Time Filter*” field in building the “*psychosis_risk*” index.
305 When exploring data in the “*Discover*” page, make sure that the index pattern “*psychosis_risk*” is
306 selected.

307 308 **4. Data visualization**

309

310 4.1 In addition to searching and accessing individual-level information via the “Discover” page in
311 Kibana, one can build visualizations and dashboards to obtain an overview of characteristics for
312 the whole population of at-risk patients. To do this, click on **Visualize** in the side navigation of
313 Kibana. Then, click the **Create new visualization** button and choose a visualization type (e.g., pie
314 and line charts). Select “*psychosis_risk*” as the index that one wants to visualize through Kibana.
315 By default, visualizations will include all records/patients in the “*psychosis_risk*” index. Detailed
316 instructions of building Kibana visualizations are available on
317 “<https://www.elastic.co/guide/en/kibana/6.4/visualize.html>”.

318

319 4.2 To select a specific subset of data for visualization, add a “*filter*”. For example, selecting a
320 filter filed as “*h_2_year*”, choosing an operator as “*is not between*” and setting values from “*0.0*”
321 to “*0.05*” will only include patients whose risk of psychosis in 2 years are higher than 0.05.

322

323 4.3 Once individual visualizations are built, click on **Dashboard** in the side navigation of Kibana to
324 create a dashboard that displays a set of related visualizations together. Click **Create new**
325 **dashboard** and the **Add** button to create a new dashboard panel. Click visualizations that one
326 wants to show within the new dashboard panel. Click **Save** and type a title to save the panel.
327 Instructions on building Kibana dashboards are available at
328 <https://www.elastic.co/guide/en/kibana/6.4/dashboard.html>.

329

330 5. Risk Alerting

331

332 5.1 Click on **Management** in the side navigation of Kibana and then click **Watcher** under
333 **Elasticsearch** to create alerting for clinicians when patients were at-risk of psychosis. If the
334 **Watcher** button is not visible, click **License Management** and click **Start trial** or **Update license**.

335

336 5.2. Click **Create advanced watch** to set up a new Watcher. Type an “ID” and “Name”. Delete the
337 content of “*Watch JSON*” section and copy the content in the “*watcher.json*” file in the
338 “*psychosis*” directory to the “*Watch JSON*” section. This watcher will send alerting Email to
339 “*clinician@nhs.uk*” (which can be replaced with the Email address where one wants to send
340 alerts) from “*username@nhs.uk*” (which was set in Step 2.3) if there are one or more patients
341 whose risk of psychosis in 2 years are higher than 0.05 (a tentative threshold for feasibility
342 testing) in every 24 hours.

343

344 5.3 Before saving the Watcher, click **Simulate** to test the Watcher execution. If the Watcher is set
345 successfully, one will see the simulation output printed. In case of error in the settings, follow the
346 instructions on [https://www.elastic.co/guide/en/elastic-stack-overview/6.4/watcher-getting-](https://www.elastic.co/guide/en/elastic-stack-overview/6.4/watcher-getting-started.html)
347 [started.html](https://www.elastic.co/guide/en/elastic-stack-overview/6.4/watcher-getting-started.html).

348

349 5.4 To stop a Watcher, permanently delete it or temporarily deactivate it from the “*Status*” page
350 of the Watcher.

351

352 REPRESENTATIVE RESULTS:

353 In this section, we present implementation results focusing in practicality in handling live clinical
354 data streams elaborated through the risk calculator and facilitating timely delivery of prognostic
355 results to clinicians. Evaluations of the clinical utility of the system, such as the adherence of
356 clinicians to the recommendations made by the risk calculator, will be presented in a separate
357 report when complete.

358

359 **Ingestion of source data**

360 We deployed the psychosis risk calculation and alerting system based on a replica database of
361 ePJS in SLaM. This replica database synchronizes the live data from ePJS every 10 minutes. A
362 database view combining patients' information for psychosis risk calculation was built in this
363 replica database, where each record contains information for a patient. All records in this view
364 were ingested into the CogStack platform in real time (approximately 0.6 microsecond per record
365 in a virtual machine with 8-core CPU and 16GB RAM). Until 13 July 2019 when this manuscript
366 was prepared, all the records of 202,289 patients who received a first index diagnosis of non-
367 organic and non-psychotic mental disorder in SLaM were ingested into CogStack for psychosis
368 risk calculation, stored in the "*psychosis_base*" Elasticsearch index. **Figure 1** shows the number
369 of records ingested into CogStack over time, in chronological order based on the last update date
370 of a record. By comparing the numbers and content of records in the database and the
371 Elasticsearch index, no missing and discrepant data were found, which confirms the reliability of
372 CogStack Pipeline in data ingestion and synchronization.

373

374 **Validation of risk results**

375 To validate the implementation of the psychosis risk detector in this protocol, we compared at-
376 risk patients detected by CogStack (called "CogStack version") with those detected by the original
377 risk calculator based on CRIS (called "CRIS version"). Since there were no thresholds developed
378 to screen an at-risk patient^{6,14,21}, we here used a tentative threshold of 5% for the risk of psychosis
379 in two years. Note that this tentative threshold is merely to test whether the system can
380 pragmatically work in the NHS and is susceptible to change with future research. The actual
381 threshold for an optimal detection of at-risk individuals will need to be identified in future large-
382 scale studies. Specifically, we first retrieved all patients who had a risk for psychosis above the
383 threshold in the CRIS version (the number of patients N=169). All these patients received a first
384 index diagnosis of non-organic and non-psychotic mental disorder in SLaM from May 14th 2018
385 to April 29th 2019. By filtering patients who were diagnosed in the same time period, we then
386 retrieved N=170 patients whose risk for psychosis in 2 years were higher than 0.05 in the
387 CogStack version. Finally, we compared the difference between the two sets of patients, where
388 the total number of unique patients in the two sets are N=173. We found that 161 patients
389 (accounting for 93% of 173 patients) had the same scores in both versions. The high degree of
390 agreement confirms the validity of this CogStack-based protocol in generating risk scores.

391

392 There were 12 patients having different risk scores in the two versions. By inspecting patients'
393 EHRs, we found that this difference was because data for these patients were updated after the
394 risk scores were calculated in the CRIS version. Specifically, although predictors used in the risk
395 calculator, such as date of birth, gender and self-assigned ethnicity, were static variables, some
396 patients' health records had a missing or default value for a variable (e.g., an unknown ethnicity)

397 at an earlier stage and these variables were entered or updated at a later stage. This can lead to
398 different risk scores at two different stages. Similarly, the first primary index diagnoses of some
399 patients were invalidated after an initial risk score was calculated based on these diagnoses. In
400 this case, the risk calculator will look for the next valid primary diagnosis for such patient and re-
401 calculate a risk score. The updated risk score can also differ from the initial one. As the original
402 risk calculator was developed based on retrospective data in CRIS for research use, the original
403 calculator pipelines did not synchronize these updates in EHR data and refresh the risk scores in
404 a timely manner. In contrast, a patient's risk score will be re-calculated in the CogStack version if
405 any source data of the patient is updated, which allows this CogStack-based calculator to provide
406 the most reliable and up-to-date risk scores for patients. These results strongly highlight the
407 reliability of risk scores in this protocol.

408

409 **Result visualization and risk alerting**

410 To demonstrate the capabilities of CogStack in data visualization, we built a dashboard for
411 information about patients at-risk of psychosis. As used before for feasibility testing, we selected
412 those who have a risk of psychosis in two years higher than 5% as at-risk patients. **Figure 2** shows
413 the visualizations of characteristics for patients at-risk of psychosis, including patients'
414 ethnicities, genders, ages and categories of diagnoses. Apart from visualizing risk results via Web
415 interfaces (e.g., Kibana), this protocol allows risk alerts to be sent to users or clinicians through
416 other notification channels such as Email. **Figure 3** shows the interface for setting a risk alerting
417 service by using the Watch component in Kibana. Once this service is configured successfully,
418 users can receive an Email notification if there were one or more patients whose risk of psychosis
419 in two years are higher than 5%. **Figure 4** shows an example of these Email notifications, which
420 reports the numbers of patients at-risk and these patients' boroughs. Since more work is needed
421 to tailor how the predicted psychosis risk scores are communicated, we have not sent risk
422 notifications directly to clinicians. For testing the technological feasibility, all notifications in this
423 study were sent from a technical researcher (T.W.) to a clinical researcher (D.O.) via the SLAM's
424 email system within a secure network. Only an aggregated statistic of patient information was
425 included in a notification; no any personally identifiable information was included.

426

427 **FIGURE AND TABLE LEGENDS:**

428 **Figure 1: Source data ingested into CogStack.** There are 202,289 records in total ingested into
429 the "*psychosis_base*" Elasticsearch index until 13 July 2019, and the histogram shows the
430 numbers of records ingested over time, ordered by the last update data time of a record. One
431 can also query both structured and unstructured information, and obtain search hits that match
432 the query in this page.

433

434 **Figure 2: Dashboard of characteristics of patients at-risk of psychosis (i.e., the risk of psychosis**
435 **in 2 years higher than 0.05).** (a) Distribution of ethnicities for patients at-risk, where outer pies
436 are the subcategories of an ethnicity category in inner pies. (b) distribution of patients' gender,
437 (c) distribution of patients' ages at diagnosis and (d) number of patients per diagnosis group.

438

439 **Figure 3: Setting and simulating Watch for risk alerting.**

440

441 **Figure 4: An example of risk alerting Email.** The numbers of patients at-risk of psychosis in each
442 Clinical Commissioning Groups (CCG) are reported in parentheses.

443

444 **DISCUSSION:**

445 We have demonstrated the first EHR implementation of a real-time psychosis risk detection and
446 alerting system based on CogStack, an open source information retrieval and extraction platform.
447 Following this approach, one can transform and ingest a large set of clinical data in various
448 formats, including structured and unstructured information, into a CogStack instance, so as to
449 enable full-text search, interactive analyses and visualization of data, as well as real-time alerting
450 to clinicians of patients that are at-risk of psychosis. Although the original psychosis risk calculator
451 has been validated in pilot studies across several NHS Trusts, albeit using retrospective patient
452 records^{6,14,21}, this experimental design provides the first evidence base that this risk calculator
453 can be replicated and deployed for use in real time. This approach allows the automatic delivery
454 of prognostic results to clinicians through existing clinical notification channels, such as Email, in
455 real time. This clearly demonstrates the technical feasibility for conducting a large-scale
456 effectiveness trial to evaluate the ultimate clinical utility of this risk calculator in the real world.

457

458 This protocol is empirically innovative, as there does not exist a similar risk detection and alerting
459 system for psychosis. Moreover, this protocol has high generalizability in clinical use, particularly
460 because of the unique strengths of our approach. From a theoretical perspective, we used a risk
461 prediction model that was developed based on a large retrospective de-identified cohort from
462 the SLaM NHS Trust. SLaM provides secondary mental health care to a population of 1.36 million
463 individuals in South London and has one of the highest recorded rates of psychosis in the world.
464 This large cohort, which has high diversity in sociodemographic and diagnostic characteristics,
465 allows us to develop a risk prediction model that is unlikely to be biased towards a population
466 with specific characteristics. This is supported by evidence that the prognostic accuracy of this
467 risk calculator has already been replicated twice in two different databases^{14,21}, including one
468 outside of SLaM. Another theoretical strength of this risk model is that basic demographic and
469 clinical diagnosis information were used as predictors. Such information is ubiquitous in
470 electronic clinical data and in fact missing data for these predictors have been shown to be
471 relatively rare in our previous studies^{14,21}. The high availability of information for building
472 predictors makes it possible to run the risk calculator over a large number of patient samples
473 across different secondary mental health care sectors. In addition, the risk calculator is a general
474 algorithm which is suitable for all individuals at-risk of developing psychosis in secondary mental
475 health care, regardless of individuals' ages. That is, this calculator is not only suitable for the 15-
476 35 age range of peak psychosis risk¹⁶, but also for those outside of this range, showing a high
477 degree of generalizability.

478

479 From a practical perspective, both the risk calculator and the CogStack platform are light-weight
480 and open-source services that do not involve resource-heavy techniques or costly infrastructure.
481 Such a low-cost and easy-to-deploy platform can reduce the barriers to its adoption in real-world
482 clinical settings. Also, our solution overcomes the main implementation barrier: risk estimation
483 systems provide little value unless they are used by clinicians in day-to-day practice²⁵. Specifically,

484 our approach accesses data from the EHR, performs analyses independent of an electronic
485 medical record system and can send analysis results back to clinicians through existing
486 notification channels. This method does not require that the business logic in pre-existing
487 systems be modified and can work as a standalone service to support and extend existing clinical
488 decision support systems. Thus, the protocol has high compatibility with pre-existing clinical
489 systems and can be easily integrated into routine clinical practice. Moreover, the protocol
490 provides user-friendly interfaces for searching, analyzing and visualizing of clinical data, which
491 make it easy for clinicians to interpret and explore the risk results.

492

493 This protocol also has its limitations. First, the effectiveness of this protocol has not been
494 evaluated in routine clinical practice. This study focused on technical feasibility tests of
495 implementing a real-time psychosis risk detection and alerting system in a local EHR. To further
496 evaluate the effectiveness of this system in routine clinical practice, future large scale
497 randomized controlled trials are needed⁶. A second limitation is that the predictions of risk scores
498 in this protocol were made based on the first primary diagnoses, which are static data collected
499 at a single snapshot. However, the CHR-P symptoms are intrinsically evolving over time. A
500 dynamic version of psychosis risk calculator, in which prediction models can be dynamically
501 updated to reflect the changes, has been developed recently²⁶. Future work will focus on
502 integrating this dynamic calculator in the current protocol.

503

504 The most critical step in this approach was identifying EHR data that were used for extract
505 predictors in the risk calculator. This may also involve creating data element mappings, when an
506 EHR system used a data model different from that used in this protocol, such as distinct coding
507 systems for patients' ethnic groups. We have open-sourced all the code and mapping definitions
508 online (<https://github.com/cogstack-slam/psychosis>). Based on these materials, one would be
509 able to replicate the deployment or adjust the calculator depending on one's own circumstance.
510 Another critical step was creating a database view for data ingestion in CogStack. Since relational
511 join operations (i.e., combining columns from one or more database tables) in Elasticsearch can
512 lead to high computational cost, we conducted these join operations in the EHR database by
513 creating a database view. This view combined all information that was needed to extract
514 predictors in the risk calculator, and two vital fields that were used by CogStack pipelines for data
515 partitioning in data ingestion. The first field is a unique primary key for each record in the view
516 ("*patient_id*" used this protocol) and the second is a timestamp when a record was modified most
517 recently. If these two fields were not set properly, CogStack might not synchronize data updates
518 in an EHR database timely. Detailed instructions for troubleshooting issues on CogStack data
519 ingestion are available on <https://cogstack.atlassian.net/wiki/spaces/COGDOC/overview> and
520 <https://github.com/CogStack/CogStack-Pipeline>.

521

522 This protocol is highly transportable and can be easily deployed in NHS Trusts that have a CRIS or
523 CogStack platform. So far, the CRIS platform—including the consenting procedures—has been
524 fully described elsewhere and is under expansion across 12 NHS Trusts in the UK, harnessing over
525 2 million deidentified patient records (<https://crisnetwork.co/>). Similarly, the CogStack platform
526 has been deployed not only in SLAM, but also other NHS Trusts across the UK such as University
527 College London Hospitals (UCLH), King's College Hospital (KCH), Guy's and St Thomas' (GSTT), and

528 Mersey Care NHS Trusts. Those Trusts without such as platform can use an online version of risk
529 calculator (<http://psychosis-risk.net>), or build this protocol from scratch based on this manuscript
530 and our online documents. Although this protocol is developed for psychosis risk detection, the
531 architectural design of this protocol is not tied-in to this specific use case. The protocol is flexible
532 enough to allow for reconfiguration and repurposing of the real-time monitoring and alerting
533 components for other risk measurement areas, such as adverse drug reactions, thereby allowing
534 clinicians to timely take action to improve patient care, safety and experience.

535

536 **ACKNOWLEDGMENTS:**

537 This study is funded by and is a direct output of the King's College London Confidence in Concept
538 award from the Medical Research Council (MRC) (MC_PC_16048) to PFP. RD and AR were
539 supported by: (a) the Maudsley Charity; (b) the National Institute for Health Research (NIHR)
540 Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's
541 College London; (c) Health Data Research UK, which is funded by the UK Medical Research
542 Council, Engineering and Physical Sciences Research Council, Economic and Social Research
543 Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish
544 Government Health and Social Care Directorates, Health and Social Care Research and
545 Development Division (Welsh Government), Public Health Agency (Northern Ireland), British
546 Heart Foundation and Wellcome Trust; (d) The BigData@Heart Consortium, funded by the
547 Innovative Medicines Initiative-2 Joint Undertaking under grant agreement No. 116074. This Joint
548 Undertaking receives support from the European Union's Horizon 2020 research and innovation
549 programme and EFPIA; it is chaired, by DE Grobbee and SD Anker, partnering with 20 academic
550 and industry partners and ESC; and (e) The National Institute for Health Research University
551 College London Hospitals Biomedical Research Centre. These funding bodies had no role in the
552 design of the study, collection and analyses. The views expressed are those of the author(s) and
553 not necessarily those of the NHS, the NIHR or the Department of Health.

554

555 **DISCLOSURES:**

556 The authors have nothing to disclose.

557

558 **REFERENCES:**

- 559 1. Lieberman, J.A., First, M.B. Psychotic disorders. *New England Journal of Medicine*. **379** (3),
560 270-280 (2018).
- 561 2. Oh, H., Koyanagi, A., Kelleher, I., DeVylder, J. Psychotic experiences and disability: findings
562 from the collaborative psychiatric epidemiology surveys. *Schizophrenia Research*. **193**, 343-347
563 (2018).
- 564 3. Tsiachristas, A., Thomas, T., Leal, J., Lennox, B.R. Economic impact of early intervention in
565 psychosis services: results from a longitudinal retrospective controlled study in England. *BMJ*
566 *Open*. **6** (10), e012611 (2016).
- 567 4. Fusar-Poli, P., McGorry, P.D., Kane, J.M. Improving outcomes of first-episode psychosis:
568 an overview. *World Psychiatry*. **16** (3), 251-265 (2017).
- 569 5. Fusar-Poli, P. The clinical high-risk state for psychosis (CHR-P), version II (2016).
- 570 6. Fusar-Poli, P. et al. Real-world Implementation of a Transdiagnostic Risk Calculator for the
571 Automatic Detection of Individuals at-risk of Psychosis in Clinical Routine: Study Protocol.

- 572 *Frontiers in Psychiatry*. **10**, 109 (2019).
- 573 7. Fusar-Poli, P. et al. Disorder, not just state of risk: meta-analysis of functioning and quality
574 of life in people at high risk of psychosis. *British Journal of Psychiatry*. **207** (3), 198-206 (2015).
- 575 8. Fusar-Poli, P. et al. Heterogeneity of psychosis risk within individuals at clinical high risk:
576 a meta-analytical stratification. *JAMA Psychiatry*. **73** (2), 113-120 (2016).
- 577 9. Fusar-Poli, P. et al. Diagnostic and prognostic significance of brief limited intermittent
578 psychotic symptoms (BLIPS) in individuals at ultra high risk. *Schizophrenia Bulletin*. **43** (1), 48-56
579 (2016).
- 580 10. Fusar-Poli, P. et al. Prognosis of brief psychotic episodes: a meta-analysis. *JAMA*
581 *Psychiatry*. **73** (3), 211-220 (2016).
- 582 11. Fusar-Poli, P., Sullivan, S., Shah, J., Uhlhaas, P. Improving the detection of individuals at
583 clinical risk for psychosis in the community, primary and secondary care: an integrated evidence-
584 based approach. *Frontiers in Psychiatry*. **10**, 774 (2019).
- 585 12. Fusar-Poli, P. Extending the benefits of indicated prevention to improve outcomes of first-
586 episode psychosis. *JAMA Psychiatry*. **74** (7), 667-668 (2017).
- 587 13. Fusar-Poli, P. et al. Transdiagnostic psychiatry: a systematic review. *World Psychiatry*. **18**
588 (2), 192-207 (2019).
- 589 14. Fusar-Poli, P. et al. Development and validation of a clinically based risk calculator for the
590 transdiagnostic prediction of psychosis. *JAMA Psychiatry*. (2017).
- 591 15. Fusar-Poli, P., Hijazi, Z., Stahl, D., Steyerberg, E.W. The science of prognosis in psychiatry:
592 a review. *JAMA Psychiatry*. **75** (12), 1289-1297 (2018).
- 593 16. Radua, J. et al. What causes psychosis? An umbrella review of risk and protective factors.
594 *World Psychiatry*. (2018).
- 595 17. Fusar-Poli, P. et al. Deconstructing vulnerability for psychosis: Meta-analysis of
596 environmental risk factors for psychosis in subjects at ultra high-risk. *European Psychiatry*. **40**,
597 65-75 (2017).
- 598 18. Fusar-Poli, P. et al. Clinical-learning versus machine-learning for transdiagnostic
599 prediction of psychosis onset in individuals at-risk. *Translational Psychiatry*. **9** (1), 1-11 (2019).
- 600 19. Stewart, R. et al. The South London and Maudsley NHS foundation trust biomedical
601 research centre (SLAM BRC) case register: development and descriptive data. *BMC Psychiatry*. **9**
602 (1), 51 (2009).
- 603 20. Jongsma, H.E. et al. Treated incidence of psychotic disorders in the multinational EU-GEI
604 study. *JAMA Psychiatry*. **75** (1), 36-46 (2018).
- 605 21. Fusar-Poli, P. et al. Transdiagnostic risk calculator for the automatic detection of
606 individuals at-risk and the prediction of psychosis: Second replication in an independent national
607 health service trust. *Schizophrenia Bulletin*. (2019).
- 608 22. Fusar-Poli, P. et al. Transdiagnostic individualized clinically based risk calculator for the
609 detection of individuals at-risk and the prediction of psychosis: Model refinement including
610 nonlinear effects of age. *Frontiers in Psychiatry*. **10**, 313 (2019).
- 611 23. Colditz, G.A., Wei, E.K. Risk prediction models: applications in cancer prevention. *Current*
612 *Epidemiology Reports*. **2** (4), 245-250 (2015).
- 613 24. Jackson, R. et al. CogStack - Experiences Of Deploying Integrated Information Retrieval
614 And Extraction Services In A Large National Health Service Foundation Trust Hospital. *BMC*
615 *Medical Informatics and Decision Making*. **18** (47) (2017).

- 616 25. McGorrian, C., Leong, T., D'Agostino, R., Graham, I. Risk estimation systems in clinical use:
617 SCORE, Heart Score and the Framingham system. *The Strateg Cardiovasc Risk 1st ed Oxford, UK*
618 *Clin Publ.* 159-172 (2008).
- 619 26. Studerus, E., Beck, K., Fusar-Poli, P., Riecher-Rössler, A. Development and Validation of a
620 Dynamic Risk Prediction Model to Forecast Psychosis Onset in Patients at Clinical High Risk.
621 *Schizophrenia Bulletin.* (2019).
- 622 27. Damschroder, L.J. et al. Fostering implementation of health services research findings into
623 practice: a consolidated framework for advancing implementation science. *Implementation*
624 *Science.* **4**, 50 (2009).
- 625 28. Fusar-Poli, P. et al. Outreach and support in south London (OASIS), 2001–2011: ten years
626 of early diagnosis and treatment for young individuals at high clinical risk for psychosis. *European*
627 *Psychiatry.* **28** (5), 315-326 (2013).
- 628