eTable 1. The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	Abstract	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Abstract Abstract NA
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Introduction		
Objectives	3	State specific objectives, including any prespecified hypotheses	Introduction and materials and methods		
Methods					
Study Design	4	Present key elements of study design early in the paper	Abstract		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Abstract, materials and methods		
Participants	6	*(a) Cohort study* - Give the	Abstract materials	RECORD 6.1: The methods of study	materials and

		eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case	and methods NA NA NA NA	population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	methods materials and methods, referenced previous publications addressing the codes or algorithms used NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	materials and methods, Table 1	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Materials and methods Table 1, Table 2, eMethods 1, eMethods 2
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	materials and methods		
Bias	9	Describe any efforts to address potential sources of bias	materials and methods		

Study size	10	Explain how the study size was arrived at	materials and methods		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Statistical analysis		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	Statistical analysis Statistical analysis and footnotes to tables Statistical analysis Study measure (time to event section) NA NA NA		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	materials and methods, external model validation NA

Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	NA
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	Results NA NA	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Results
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	Table 3 Table 3 Results		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure	Results and Figure 1 NA		

		Cross-sectional study - Report numbers of outcome events or summary measures	NA		
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Results NA Figure 1		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	Results		
Discussion					
Key results	18	Summarise key results with reference to study objectives	First paragraph of discussion		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Discussion (strengths and weaknesses of the study section)	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Discussion (implications for clinical practice and strengths and weaknesses of the study sections)
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar	Discussion (interpretation of findings and strengths and weaknesses of the		

		studies, and other relevant evidence	study sections)		
Generalisability	21	Discuss the generalisability (external validity) of the study results	Discussion		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Financial support statement in the abstract		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Statistical analysis, eMethods 2

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

eMethods 1. Definition of emerging ICD-10 primary diagnosis of non-organic psychotic disorder

This was defined as the emergence of the first ICD-10[1] primary diagnosis of non-organic psychotic disorder, occurring at least three months after the index diagnosis as recorded in the local electronic medical records: schizophrenia spectrum psychoses (schizophrenia [F20.x, except F20.4/F20.5], schizoaffective disorder [F25.x], delusional disorders [F22.x, F24], acute and transient psychotic disorders [F23.x]), unspecified nonorganic psychosis (F28/F29), psychotic disorders due to psychoactive substance use ([F10-F19].5), and affective psychoses (mania with psychotic symptoms [F30.2], bipolar affective disorder with psychotic symptoms [F31.2, F31.5], and depression with psychotic symptoms [F32.3/F33.3]). Accordingly, baseline ICD-10 psychotic disorders were excluded, with the exception of Acute and Transient Psychotic Disorders (ATPD, F23.x), which are, by definition, clinically remitting and non-psychotic within three months (short-lived). The rationale for including the ATPD is due to the fact that this group is prognostically similar to the Brief Limited Intermittent Psychotic Symptom (BLIPS) or Brief Limited Psychotic Symptoms (BIPS) subgroups of the CHR-P construct (for details on these competing operationalization see previous publications on the diagnostic and prognostic significance of BLIPS[2, 3]). On a diagnostic level, about two thirds (68%) of BLIPS meet ATPD criteria [2].

eTable 2. Predictor definitions: Primary index diagnoses of non-organic and non-psychotic mental disorder formulated at baseline (time of the first contact with the NHS Trust).

Primary index diagnosis	ICD-10 code	ICD-10 diagnosis name
Acute and transient psychotic disorders	F23.x	Acute and transient psychotic disorders
Substance use disorders	F10 (excluding *.5, *.4, *.7)	Non psychotic mental and behavioural disorders due to use of alcohol
	F11 (excluding *.5, *.4, *.7)	Non psychotic mental and behavioural disorders due to use of opioids
	F12 (excluding *.5, *.4, *.7)	Non psychotic mental and behavioural disorders due to use of cannabinoids
	F13 (excluding *.5, *.4, *.7)	Non psychotic mental and behavioural disorders due to use of sedatives or hypnotics
	F14 (excluding *.5, *.4, *.7)	Non psychotic mental and behavioural disorders due to use of cocaine
	F15 (excluding *.5, *.4, *.7)	Non psychotic mental and behavioural disorders due to use of other stimulants, including caffeine
	F16 (excluding *.5, *.4, *.7)	Non psychotic mental and behavioural disorders due to use of hallucinogens
	F17 (excluding *.5, *.4, *.7)	Non psychotic mental and behavioural disorders due to use of tobacco
	F18 (excluding *.5, *.4, *.7)	Non psychotic mental and behavioural disorders due to use of volatile solvents
	F19 (excluding *.5, *.4, *.7)	Non psychotic mental and behavioural disorders due to multiple drug use and use of other psychoactive substances
Bipolar mood disorders	F31.x (excluding F31.2 and F31.5)	Non psychotic bipolar disorder
	F34.0	Cyclothymia
	F30.x (excluding *.2)	Non psychotic mania or hypomania
Non bipolar mood disorders	[F32-F33].x (excluding F32.3 and F33.3)	Non psychotic depressive disorder
	F34.1	Dysthymia
	F34.8, F34.9, F38.x, F39	Unspecified mood disorders
Anxiety disorders	F40.x	Phobic anxiety disorders
	F41.0	Panic disorder
	F41.1	Generalized anxiety disorder
	F41.2-F41.9	Other anxiety disorders
	F42.x	Obsessive compulsive disorders
	F43.x	Reaction to severe stress, and adjustment disorders
	F44.x	Dissociative [conversion] disorders
	F45.x	Somatoform disorders
	F48.x	Other neurotic disorders
Personality disorders	F60.0	Paranoid personality disorder
	F60.1	Schizoid personality disorder
	F60.2	Dissocial personality disorder
	F60.3	Emotionally unstable personality disorder
	F60.4	Histrionic personality disorder
	F60.5	Anankastic personality disorder

Primary index diagnosis	ICD-10 code	ICD-10 diagnosis name
	F60.6	Anxious [avoidant] personality disorder
	F60.7	Dependent personality disorder
	F60.8-F60.9, F61, F62.x, F68.x, F69	Other personality disorders
	F21	Schizotypal Disorder
	F63.x	Habit and impulse disorders
	F64.x, F65.x, F66.x	Sexual disorders
Developmental disorders	F80.x	Specific developmental disorders of speech and language
	F81.x, F82, F83	Other specific developmental disorders
	F84.x	Pervasive developmental disorders
	F88, F89	Other and unspecified disorders of psychological development
Childhood/adolescence onset disorders	F90.x	Hyperkinetic disorders
	F91.x	Conduct disorders
	F92.x, F93.x, F94.x, F98.x	Other emotional and behavioural disorders with childhood or adolescence onset
	F95.x	Tic disorders
Physiological syndromes	F50.x	Eating disorders
	F51.x	Nonorganic sleep disorders
	F52.x	Sexual dysfunction, not caused by organic disorder or disease
	F53.x (excluding F53.1)	Non psychotic Mental and behavioural disorders associated with the puerperium, not elsewhere classified
	F54.x, F55, F59	Other physiological syndromes
Mental retardation	F70.x	Mild mental retardation
	F71.x	Moderate mental retardation
	F72.x	Severe mental retardation
	F73.x	Profound mental retardation
	F78.x	Other mental retardation
	F79.x	Unspecified mental retardation
F00-F09 organic mental disorders and all psychotic disorders other than F23.x were excluded		

eTable 3. Predictor definitions: self-assigned ethnicity.

Ethnic group	**Ethnicity as recorded in patient electronic health records**
Black	Black or Black British - African Black or Black British - Caribbean Black or Black British - Any other Black background
White	White - British White - Irish White - Any other White background
Asian	Asian or Asian British - Bangladeshi Asian or Asian British - Indian Asian or Asian British - Pakistani Asian or Asian British - Any other Asian background Other Ethnic Groups - Chinese
Mixed	Mixed - White and Asian Mixed - White and Black African Mixed - White and Black Caribbean Mixed - Any other mixed background
Other	Other Ethnic Groups - Any other ethnic group
Missing	Not Known Not Recorded

eMethods 2. STATA scripts used for the external model validation in the C&I NHS Trust

**** Validation of the SLAM model in the C&I trust ****

```
generate xb_d=(0.5681779*i1.gender+0.0117113*age_diag-
0.0121931*age_diag*i1.gender+1.037915*i2.ethnicity_gr+0.5143438*i3.ethnicity_g
r+0.6044039*i4.ethnicity_gr+0.4081036*i5.ethnicity_gr+0.9867204*i1.base_dgs-
1.925903*i2.base_dgs-0.1754082*i3.base_dgs-1.886428*i4.base_dgs-
2.235825*i5.base_dgs-1.547794*i6.base_dgs-3.466732*i7.base_dgs-
3.25382*i8.base_dgs-2.463145*i9.base_dgs-2.450679*i10.base_dgs)
summarize xb_d
stcox xb_d, nohr basesurv(surv0)
predict hr
estat concordance
generate surv1= surv0^exp(xb_d)
sum surv0 if _t<3650
scalar base10y=r(min)
generate surv10y_d=1-base10y^exp(xb_d)
brier psychosis_10yrs surv10y_d
```

**** Model recalibration ****

```
generate xb_d_calibrated=xb_d* .7502673
summarize xb_d_calibrated
stcox xb_d_calibrated, nohr basesurv(surv0)
predict hr
estat concordance
generate surv1= surv0^exp(xb_d_calibrated)
sum surv0 if _t<3650
scalar base10y=r(min)
generate surv10y_d=1-base10y^exp(xb_d_calibrated)
brier psychosis_10yrs surv10y_d
```

eTable 4 Transdiagnostic risk calculator in the C&I dataset compared to the original model as developed in the SLaM dataset (Lambeth and Southwark)

Variable	C&I		SLaM (Lambeth and Southwark)	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Age, y	0.988 (0.978-0.998)	.016	1.011 (1.007-1.017)	<.001
Sex				
Male	3.090 (1.815-5.261)	<.001	1.764 (1.298-2.399)	<.001
Female	1			
Race/ethnicity				
White	1		1	
Black	2.332 (1.851-2.938)	<.001	2.823 (2.439- 3.268)	<.001
Asian	1.211 (0.838-1.752)	.308	1.673 (1.216- 2.30)	.002
Mixed	1.836 (1.239-2.720)	.002	1.830 (1.276-2.625)	.001
Other	1.241 (0.872-1.768)	.230	1.504 (1.210-1.869)	<.001
Index diagnosis				
Acute and transient psychotic disorders	9.728 (7.249-13.053)	<.001	17.693 (14.334-21.839)	<.001
Substance use disorders	1.073 (0.796-1.448)	0.644	0.961 (0.762-1.213)	.740
Bipolar mood disorders	2.722 (1.971-3.758)	<.001	5.535 (4.336-7.065)	<.001
Nonbipolar mood disorders	1		1	
Anxiety disorders	0.804 (0.574-1.125)	.203	0.705 (0.560- 0.888)	0.003
Personality disorders	1.250 (0.865-1.807)	.234	1.403 (0.997- 1.974)	0.052
Developmental disorders	1.637 (0.596-4.500)	.339	0.206 (0.103-0.411)	<.001
Childhood/adolescence onset disorders	-	-	0.255 (0.169-0.385)	<.001
Physiological syndromes	0.863 (0.211-3.526)	.837	0.562 (0.369-0.856)	.007
Mental retardation	3.721 (1.504-9.207)	.004	0.569 (0.339-0.955)	.033
CHR-P	-		6.596 (4.752-9.155)	<.001
Age * sex	1.025 (1.012-1.038)	<.001	0.988 (0.981-0.995)	.001

The follow-up in C&I was truncated at 6 years for comparability purposes. For the same reason the reference group of the index diagnosis in the SLaM model was changed from CHR-P to nonbipolar mood disorders.

eLimitations

We did not employ structured psychometric interviews to ascertain the type of emerging psychotic diagnoses at follow up. However, we predicted psychotic disorders rather than specific ICD-10 diagnoses, a category which has good prognostic stability[59]. Therefore, while the psychotic diagnoses in our analyses are high in ecological validity (i.e. they represent real-world clinical practice), they have not been subjected to formal validation with research-based criteria. However, the use of structured diagnostic interviews can lead to selection biases, decreasing the transportability of models[4]. There is also meta-analytical evidence indicating that within psychotic disorders, administrative data recorded in clinical registers are generally predictive of true validated diagnoses[60]. Another limitation is that the research team carrying out this replication study is not completely independent from the original research team that developed the model[61]. However, independent external validation by completely different teams in biomedical research remains rare[39]. To facilitate further replication studies, we have provided details for operationalizing the predictors and outcomes entered in the model, and we have appended the statistical scripts used in the analyses. It is also possible that the model is charting out relationships that reflect diagnostic practice within the UK health registry system: replication studies outside the UK are needed to clarify this. Finally, although we welcome further external validation studies, it must be noted that even strong replication does not automatically imply the potential for successful adoption in clinical or public health practice. Ideally, randomized clinical trials or economic modelling are needed to assess whether our risk calculator effectively improves patient outcomes.

REFERENCES

1. WHO. *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. Geneva, ; 1992.

2. Fusar-Poli P, Cappucciati M, De Micheli A, et al. Diagnostic and Prognostic Significance of Brief Limited Intermittent Psychotic Symptoms (BLIPS) in Individuals at Ultra High Risk. *Schizophr Bull* Jan 2017;43(1):48-56.

3. Fusar-Poli P, Cappucciati M, Bonoldi I, et al. Prognosis of brief psychotic episodes: a meta-analysis. *JAMA Psychiatry* 2016;73(3):211-220.