BJPsych | The British Journal of Psychiatry (2025)
1–8. doi: 10.1192/bjp.2025.10313

Original Article

Prediction of individualised 6-month mortality risk in opioid use disorder

Emmert Roberts, John Strang, Eve Taylor, Jamie Crummy, Tim Lowden, Chioma Amasiatu and Brian Eastwood

Background

People with opioid use disorder (OUD) have substantially higher standardised mortality rates compared with the general population. However, lack of individualised prognostic information presents challenges in personalisation of addiction treatment delivery.

Aims

To develop and validate the first prognostic models to estimate 6-month all-cause and drug-related mortality risk for people diagnosed with OUD using indicators recorded at baseline assessment in addiction services in England.

Method

Thirteen candidate prognostic variables, including sociodemographic, injecting status and health and mental health factors, were identified from nationally linked addiction treatment, hospital admission and death records from 1 April 2013 to 1 April 2022. Multivariable Cox regression models were developed with a fractional polynomial approach for continuous variables, and missing data were addressed using multiple imputation by chained equations. Validation was undertaken using bootstrapping methods. Discrimination was assessed using Harrel's C and D statistics alongside examination of observed-to-predicted event rates and calibration curve slopes.

Results

Data were available for 236 064 people with OUD, with 2427 deaths due to any cause, including 1289 due to drug-related causes. Both final models demonstrated good optimism-adjusted discrimination and calibration, with all-cause and drug-related models, respectively, demonstrating Harrell's C statistics of 0.73 (95% CI 0.71–0.75) and 0.74 (95% CI 0.72–0.76), D-statistics of 1.01 (95% CI 0.95–1.08) and 1.07 (95% CI 0.98–1.16) and calibration slopes of 1.01 (95% CI 0.95–1.08) and 1.01 (95% CI 0.94–1.10).

Conclusions

We developed and internally validated Roberts' OUD mortality risk, with the first models to accurately quantify individualised absolute 6-month mortality risks in people with OUD presenting to addiction services. Independent validation is warranted to ensure these models have the optimal utility to assist wider future policy, commissioning and clinical decision-making.

Keywords

Mortality; risk prediction; opioid use disorder; prognosis; epidemiology.

Copyright and usage

© The Author(s), 2025. Published by Cambridge University Press on behalf of Royal College of Psychiatrists. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

In 2022, England reported its highest number of drug-related deaths on record.[1] Opioids were implicated in almost half of all drug-related deaths, and opioid use disorder (OUD) was a problem for half of all adults accessing community addiction services.[2] Over the past 8 years, between 1 and 2% of all adults accessing community addiction services with OUD have died each year while receiving treatment.[2] Professionals working in community addiction services have key roles in both the delivery of evidence-based treatment and the provision of prognostic information to individuals with OUD. However, despite a good understanding that people with OUD have, on average, standardised mortality rates up to ten times higher than those of the general population,[3,4] uncertainty regarding individual prognosis and mortality risk presents challenges to addiction services in provision of individuals with accurate personalised risk information, prioritisation of finite resources and appropriate targeting of interventions.

Expansion in the use of clinical informatics and precision medicine has revolutionised the care provided in many healthcare sectors;[5] however, development and validation of prognostic risk models in populations of people with OUD has been relatively limited. This is despite multiple systematic reviews having examined individual prognostic risk factors for mortality among people with OUD[3,4,6–8] and several studies having recently developed models in populations routinely prescribed opioids (e.g. to examine the risk of developing OUD or the risk of opioid overdose).[9,10] To our knowledge, no models have been developed to examine mortality risk in people with a diagnosis of OUD presenting to community addiction services. Models that examine both all-cause and drug-related mortality risks could provide useful information and assistance to both individuals and professionals in making collaborative treatment decisions at the clinically important point of entering addiction treatment.

Potential explanations for the relative paucity of prognostic modelling studies in this area include the required sample size and number of events, and a lack of centralised data repositories that include linked accurate prognostic and outcome information from healthcare and administrative agencies. England is unusual, having recently established and validated a 10-year national data linkage between all individuals presenting to community addiction services and their hospital admission and death records ($n > 900\,000$).[11] This is coupled with the fact that all people in England, regardless of overseas visitor or immigration status, are able to access community addiction services free of charge at the point of delivery; hence, there is a relative absence of a privately funded treatment system.[12] The availability and coverage of this national linked data-set thus provides a rare opportunity to develop and validate adequately powered prognostic models within this population.

https://doi.org/10.1192/bjp.2025.10313 Published online by Cambridge University Press

Objectives

In this study, we aimed to develop and validate two models, one to estimate 6-month all-cause mortality risk and the other to estimate 6-month drug-related mortality risk for people with OUD on the basis of prognostic indicators routinely recorded during initial assessment at community addiction services in England.

Method

The complete protocol has been previously published,[13] and the study was designed and reported in accordance with the Transparent Reporting of multivariable prediction models for Individual Prognosis Or Diagnosis that use regression or machine learning methods (TRIPOD + AI) statement.[14] The completed TRIPOD + AI checklist is provided as supplementary Table 1 available at https://doi.org/10.1192/bjp.2025.10313. The work benefited throughout from input from the South London and the Maudsley Biomedical Research Centre Data Linkage Service User and Carer Advisory Group, which includes experts with lived experience of OUD.[15]

Setting

In this study, we used a national English data-set containing linked individual records from three sources: (a) the National Drug Treatment Monitoring System (NDTMS), (b) Hospital Episode Statistics (HES) and (c) the Office of National Statistics (ONS). NDTMS is a centralised database collated and maintained by the Department of Health and Social Care (DHSC) that receives monthly input from all adult statutory community addiction services in England[16] and contains individual-level data on sociodemographic characteristics (date of birth, sex, housing status, etc.), what substances the individual is using problematically, any interventions received and measures of treatment success. HES is a centralised database collated and maintained by the National Health Service (NHS) that collects all information pertaining to NHS in-patient hospital admission in England[17] and covers all NHS in-patient admissions, including any admissions to private or third-sector hospitals subsequently reimbursed by the NHS; it is estimated to contain >99% of all in-patient hospital activity in England. An in-patient hospital admission includes any secondary-care-based activity requiring a hospital bed; this includes day cases and both planned and emergency admissions in physical and mental health settings. HES does not cover accident and emergency (emergency department) attendances, nor out-patient bookings, as these data are held in separate databases. The ONS is a centralised database that contains official death certification records for those individuals that have died. The overall structure of the linked NDTMS-HES-ONS data is clustered with individuals attending one of 150 uniquely commissioned statutory community addiction services across each local authority area in England.

Approval to conduct the linkage analyses was granted under regulation 3 of the Health Service (Control of Patient Information) Regulations 2002, following review by the Caldicott Advisory Panel (ref: CAP-2019-06) and the DHSC Office of Data Protection. NDTMS data were available from 1 April 2013 to 1 April 2022 and contained information for 236 064 unique individuals aged 18 years or over who had attended community addiction services for treatment of problematic opioid use at least once within that timeframe. A standardised clinical history supported by urine drug screen investigation is used within community addiction services to confirm and record a diagnosis of OUD. Linked HES and ONS data are available for individuals; these data include details of any subsequent death records and any individual hospital admissions since HES database inception in 1997.[11] The linked database can only be accessed by DHSC staff working on the project, and all records are stored for a minimum of 5 years after study completion. The studied timeframe deviated from the published protocol, as the end date of the study window was previously specified as 1 April 2023. This deviation was necessary because accurate death record outcome data were not yet available for all individuals presenting to addiction services after 1 April 2022.

Candidate predictor variables

The prognostic indicators for consideration in the multivariable model were initially identified from a systematic search of review articles and their included studies which examined demographic and clinical features associated with increased mortality for people with OUD.[3,4,6–8,18] All candidate predictor variables that were significantly associated with all-cause mortality were extracted, and those which were available within the linked NDTMS-HES-ONS records were discussed with patients and clinicians over a series of three consensus meetings. Given that the aim was to create a model that could be readily incorporated into routine clinical care within time-pressured services, a parsimonious approach was taken to select prognostic indicators, with clinician and patient involvement suggesting that ideally no more than ten variables should be included in the final model. An agreed consensus set of prognostic indicator variables was subsequently extracted from NDTMS-HES-ONS records retrospectively from the time of baseline assessment (i.e. the point at which the patient initially presented for treatment) at the community addiction service, designated time zero (t_0). Following discussion, the protocol initially identified 12 candidate prognostic indicator variables; however, on the basis of additional patient and clinician input during model development, a supplementary candidate variable, the binary of whether a person had ever previously been in addiction treatment, was added as a candidate predictor. Descriptions and structures of the candidate variable can be found in Table 1.[19,20]

Outcome measures

The binary outcomes of all-cause and drug-related mortality were assessed prospectively for each individual at any point up to 6 months after t_0. This timepoint was chosen following clinician and patient feedback, as it was thought to reflect a time horizon of sufficient duration to potentially encourage risk factor modification. For drug-related death, we followed the definition used by the ONS when reporting official national statistics for deaths related to drug-poisoning. The death certificate ICD-10 codes for drug-related deaths can be found in supplementary Table 2; these include codes for deaths due to mental and behavioural disorders caused by drug use (ICD-10: F11–F16, F18–F19), assault (ICD-10: X85), and poisoning of accidental, intentional and undetermined intent (ICD-10: X40–X44, X60–X64 and Y10–Y14).

Sample size

The minimum required sample size for time-to-event model development was based on estimated event rates of prediction model outcomes.[21] Given that the drug-related death event rate was by definition smaller than the all-cause death rate and thus required a larger sample size, this outcome was chosen for sample size calculation. Estimation was performed using the 'pmsampsize' command, and, owing to the absence of any reported Cox–Snell R-squared values from previously developed models, we aimed to develop a model with a minimal anticipated Harrel's C statistic (a measure of discrimination that is similar to the area under a

https://doi.org/10.1192/bjp.2025.10313 Published online by Cambridge University Press

Table 1 Candidate predictor variables

Candidate predictor variable	Variable structure in NDTMS-HES-ONS
Age	Continuous
Sex	Binary: 0: Female 1: Male
History of injecting behaviour	Categorical: 0: Never injected 1: Previously injected (but not currently) 2: Currently injecting
HIV positivity[a]	Binary: 0: No 1: Yes
Hepatitis C RNA status[a]	Binary: 0: Negative (never infected or cleared by treatment) 1: Positive
Polysubstance use: Number of substances used problematically[b]	Categorical: 0: One problematic substance 1: Two problematic substances 2: Three or more problematic substances
Problematic alcohol use[b]	Binary: 0: No problematic use of alcohol 1: Problematic use of alcohol
Problematic benzodiazepine use[b]	Binary: 0: No problematic use of any benzodiazepine 1: Problematic use of any benzodiazepine
Accommodation need	Categorical: 0: No housing problem - Owner occpier - Tenant – private landlord, housing association, local authority, registered landlord or arm's length management organisation - Approved premises - Supported housing or hostel - Traveller - Own property - Settled mainstream housing with friends and/or family - Shared ownership scheme 1: Housing problem, i.e. - staying with friends and/or family as a short-term guest - night winter shelter - direct access short-stay hostel - short-term B&B or other hotel - placed in temporary accommodation by local authority - squatting 2: No fixed abode – urgent housing problem, i.e. - lives on streets or rough sleeper - uses night shelter (night-by-night basis) or emergency hostels - sofa surfing or sleeps on different friend's floor each night
Prison referral	Binary: 0: Referred to the drug service by any source other than prison 1: Referred to the drug service from prison
Acute in-patient hospital admission[c]	Binary: 0: No acute in-patient hospital admissions within the person's lifetime 1: Acute in-patient hospital admission within the person's lifetime
Mental health in-patient hospital admission[c]	Binary: 0: No in-patient involuntary mental health hospital detention within the person's lifetime 1: In-patient involuntary mental health hospital detention within the person's lifetime
Previous history of addiction treatment[d]	Binary: 0: The person has never previously had an episode of addiction treatment 1: The person has previously had an episode of addiction treatment

NDTMS-HES-ONS, National Drug Treatment Monitoring System, Hospital Episode Statistics and Office of National Statistics.
a. Available from 2020 onwards.
b. Problematic use as deemed by the assessing clinician.
c. Variation from protocol which specified that the acute and mental health in-patient hospital admissions be within the past 6 months; following continued patient and clinician input during model development, this was revised to lifetime admission to hospital, as this was thought to be an easier question to ask individuals with opiod use disorder, and the strict timeframe could have led to implementation issues with checking records.
d. Variation from protocol: this predictor was added following patient and clinician input during model development.

receiver operating characteristic curve but takes account of the censored nature of the data) of 0.70, allowing a maximum shrinkage of 10% to minimise potential overfitting.[22] A maximum total of 12 candidate predictors was originally planned, with an estimated event rate based on a previous cohort study which reported 0.0134 drug-related deaths per person year.[18] The estimated minimum required sample size was 2487 participants and 51 events.

Missing data

Complete outcome data were available for all individuals, and complete candidate predictor information was available for eight variables. Of the five candidate predictors with missing data (injecting status, HIV positivity, hepatitis C RNA positivity, prison referral and accommodation need) the fraction of missing information and its assumed missingness mechanism were assessed for each variable, and all were deemed to at least reasonably fulfil the missing at random assumption (i.e. the probability of a value's being missing in one variable was not deemed to be related to the probability of missing data in another variable). Missing data were addressed using multiple imputation by chained equations, with the number of imputations set to $m = 50$ on the basis of the highest fraction of missing information.[23] Rubin's rules were used to combine the results across imputed data-sets.[24]

Statistical analysis

Multivariable Cox regression models were developed using backwards elimination with the level of alpha for variable exclusion set to 0.157, as recommended on the basis of the Akaike information criterion.[25,26] Non-linearity of continuous variables was addressed by a multivariable fractional polynomial approach, an established technique for transforming non-linear continuous variables when developing a backwards elimination model.[27] Model discrimination was assessed through calculation of Harrel's C statistic and D statistic (a measure of discrimination, with higher values indicating better discrimination), and we examined calibration curve slopes and the ratio of observed to predicted event rates.[27,28] Validation was undertaken using bootstrapping resampling methods, which account for bias due to overfitting more accurately than split sample cross-validation approaches. The model development process was repeated in 1000 bootstrap

https://doi.org/10.1192/bjp.2025.10313 Published online by Cambridge University Press

Table 2 Baseline characteristics of people with opioid use disorder presenting to community addiction services in England between 1 April 2013 and 1 April 2022

	Full sample, *n*	Died due to any cause within 6 months, *n*	Died due to a drug-related cause within 6 months, *n*
All	236 064 (100%)	2427 (100%)	1289 (100%)
Mean (s.d.) age in years	43.4 (9.2)	46.2 (9.6)	43.8 (8.6)
Sex			
Female	63 443 (26.9%)	538 (22.2%)	291 (22.6%)
Male	172 621 (73.1%)	1889 (77.8%)	998 (77.4%)
History of injecting behaviour			
Never injected	98 542 (41.7%)	565 (23.3%)	229 (17.8%)
Previously injected (but not currently)	72 085 (30.5%)	944 (38.9%)	517 (40.1%)
Currently injecting	63 542 (26.9%)	892 (36.8%)	532 (41.3%)
Missing[1]	1895 (0.8%)	26 (1.1%)	11 (0.9%)
HIV status			
Negative	97 912 (41.5%)	271 (11.2%)	140 (10.9%)
Positive	2132 (0.9%)	11 (0.5%)	8 (0.6%)
Missing[1]	136 020 (57.6%)	2145 (88.4%)	1141 (88.5%)
Hepatitis C RNA status			
Negative	86 931 (36.8%)	417 (17.2%)	233 (18.1%)
Positive	10 443 (4.4%)	148 (6.1%)	79 (6.1%)
Missing[1]	138 690 (58.8%)	1862 (76.7%)	977 (75.8%)
Polysubstance use: number of substances used problematically			
1	63 000 (26.7%)	565 (23.3%)	277 (21.5%)
2	99 822 (42.3%)	997 (41.1%)	505 (39.2%)
≥3	73 242 (31.0%)	865 (35.6%))	507 (39.3%)
Problematic alcohol use			
No	203 307 (86.1%)	1875 (77.3%)	1014 (78.7%)
Yes	32 757 (13.9%)	552 (22.7%)	275 (21.3%)
Problematic benzodiazepine use			
No	219 926 (93.2%)	2189 (90.2%)	1113 (86.3%)
Yes	16 138 (6.8%)	238 (9.8%)	176 (13.7%)
Accommodation need			
No housing problem	153 346 (64.7%)	1595 (65.7%)	829 (64.3%)
Housing problem	31 677 (13.4%)	404 (16.7%)	221 (17.2%)
Urgent housing problem – no fixed abode	25 370 (10.8%)	364 (15.0%)	217 (16.8%)
Missing[a]	25 671 (10.9%)	64 (2.6%)	22 (1.7%)
Prison referral			
Not referred from prison	203 012 (86.0%)	1983 (81.7%)	1012 (78.5%)
Referred from prison	31 774 (13.5%)	427 (17.6%)	268 (20.8%)
Missing[a]	1278 (0.5%)	17 (0.7%)	9 (0.7%)
Acute inpatient hospital admission			
None	76 942 (32.6%)	652 (26.9%)	375 (29.1%)
Previous admission	159 122 (67.4%)	1775 (73.1%)	914 (70.9%)
Mental health inpatient hospital admission			
None	235 499 (99.8%)	2412 (99.4%)	1280 (99.3%)
Previous admission	565 (0.2%)	15 (0.6%)	9 (0.7%)
Previous history of addiction treatment			
First treatment episode	83 222 (35.3%)	469 (19.3%)	248 (19.2%)
Previous treatment episode	152 842 (64.7%)	1958 (80.7%)	1041 (80.8%)

a. Proportion of missing values imputed as variable values can be found in the online supplementary material in Table S4 in the online supplementary material.

samples to allow calculation of optimism-adjusted discrimination and calibration measures.[29] Performance was also evaluated by calculation of Harrell's C statistics for each cluster (i.e. each of the 150 individual statutory community addiction services), and the results were combined using random effects meta-analysis, weighted by the number of events per service. Between-cluster heterogeneity was assessed using the I^2 statistic.[28] All analyses were conducted in Stata version 18.0 (StataCorp, College Station, Texas, USA).

Results

Data were available for 236 064 people with OUD. There were 2427 deaths due to any cause and 1289 deaths due to a drug-related cause within 6 months of the individual's most recent presentation to community addiction services in England. Baseline characteristics of the whole sample and those dying due to any or drug-related causes within 6 months are available in Table 2.

Development

Table 3 shows the optimism-adjusted hazard ratios (aHRs) for both final models. In both models, the final model and all variables met the assumption of proportional hazards.

All-cause mortality

All variables were preserved in the all-cause mortality model except HIV positivity and polysubstance use, leading to a final model event per variable (EPV) rate of 2427/13 = 186.69. Supplementary Fig. 1 shows a graphical representation of the aHRs for the fractional polynomial terms for age; this was treated as a cubic function in the final model. For the variables included in the final model, the

https://doi.org/10.1192/bjp.2025.10313 Published online by Cambridge University Press

Table 3 Optimism-adjusted hazard ratios (95% CIs) for 6-month all-cause and drug-related mortality in individuals with opioid use disorder presenting to community addiction services in England

	All-cause mortality	Drug-related mortality
Age, years[a]	1.00 (1.00–1.00)	N/A
Sex		
Female	Reference	Reference
Male	1.15 (1.04–1.26)	1.16 (1.01–1.32)
History of injecting behaviour		
Never injected	Reference	Reference
Previously injected (but not currently)	1.90 (1.71–2.12)	2.57 (2.19–3.02)
Currently injecting	2.06 (1.84–2.30)	2.87 (2.44–3.38)
Hepatitis C RNA status		
Negative	Reference	Reference
Positive	1.29 (1.12–1.49)	1.26 (1.03–1.53)
Problematic alcohol use		
No	Reference	Reference
Yes	1.73 (1.57–1.91)	1.67 (1.46–1.91)
Problematic benzodiazepine use		
No	Reference	Reference
Yes	1.34 (1.17–1.53)	1.86 (1.58–2.18)
Accommodation need		
No housing problem	Reference	Reference
Housing problem	1.13 (1.02–1.26)	1.12 (0.96–1.30)
Urgent housing problem – no fixed abode	1.17 (1.04–1.32)	1.20 (1.03–1.41)
Prison referral		
Not referred from prison	Reference	Reference
Referred from prison	1.35 (1.21–1.50)	1.46 (1.27–1.68)
Acute in-patient hospital admission		
None	Reference	Reference
Previous admission	1.18 (1.08–1.30)	1.13 (1.00–1.27)
Mental health in-patient hospital admission		
None	Reference	Reference
Previous admission	2.23 (1.34–3.71)	2.67 (1.39–5.16)
Previous history of addiction treatment		
First treatment episode	Reference	Reference
Previous treatment episode	1.93 (1.74–2.12)	1.89 (1.64–2.18)
Baseline survivor function[b]	0.9917	0.9958
Mean linear predictor (s.d., range)	1.62 (0.66, 0.03–5.17)	1.40 (0.73, 0.00–4.30)

a. Modelled as a cubic function in all-cause mortality model.
b. Continuous covariates set at their means and categorical or binary variables set at their reference values.

Table 4 Mean (95% CI) performance of 6-month all-cause and drug-related mortality prediction models in people with opioid use disorder presenting to community addiction services in England

	All-cause mortality		Drug-related mortality	
	Original apparent	Optimism adjusted	Original apparent	Optimism adjusted
Harrell's C statistic	0.73 (0.71–0.75)	0.73 (0.71–0.75)	0.74 (0.72–0.76)	0.74 (0.72–0.76)
D Statistic	1.02 (0.96–1.07)	1.01 (0.95–1.08)	1.09 (1.00–1.17)	1.07 (0.98–1.16)
R^2_D	0.20 (0.18–0.22)	0.20 (0.18–0.22)	0.22 (0.19–0.25)	0.21 (0.19–0.24)
Calibration slope	1 (0.94–1.06)	1.01 (0.95–1.08)	1 (0.92–1.08)	1.01 (0.94–1.10)

optimism-adjusted increase in individual risk of all-cause mortality ranged between a 13% increase (95% CI 2–26%) for having a non-urgent housing problem to a 123% increase (95% CI 34–271%) for having a previous involuntary mental health admission.

Drug-related mortality

All variables were preserved in the drug-related mortality model except HIV positivity, polysubstance use and age, for which fractional polynomial terms did not reach the prespecified significance for inclusion at either the first or the second degree. This led to a final model EPV rate of 1289/12 = 107.42. For the variables included in the final model, the optimism-adjusted increase in individual risk of drug-related mortality ranged between a 12% increase (95% CI 0–27%) for having a previous acute hospital admission to a 187% increase (95% CI 144–238%) for being a person who currently injects.

Supplementary Table 3 shows the complete case analysis (i.e. the results based only on people with complete candidate predictor variable data); the aHRs showed broadly similar trends in both final models using the multiply imputed data.

Validation

Table 4 shows the optimism-adjusted performance statistics for both final models.

Discrimination

All-cause mortality

The optimism-adjusted final model explained 20% of the variation in time to all-cause mortality (R^2_D), the D statistic was 1.01 and Harrell's C statistic was 0.73. Supplementary Fig. 2 shows a forest plot of Harrel's C statistics across individual community addiction

https://doi.org/10.1192/bjp.2025.10313 Published online by Cambridge University Press

Table 5 Clinical examples of all-cause and drug-related 6-month mortality risk for people with opioid use disorder presenting to community addiction services in England

	Example			
	1	2	3	4
Age,years	18	52	66	53
Sex	Female	Male	Male	Male
History of injecting behaviour	Never injected	Previously injected	Currently injecting	Currently injecting
Hepatitis C RNA status	Negative	Negative	Negative	Negative
Problematic alcohol use	No	Yes	Yes	Yes
Problematic benzodiazepine use	No	No	No	Yes
Accommodation need	No housing problem	No housing problem	Housing problem	Housing Problem
Prison referral	No	No	No	No
Acute in-patient hospital admission	No	No	Yes	Yes
Mental health in-patient hospital admission	No	No	No	Yes
Previous history of addiction treatment	No	Yes	Yes	Yes
All-cause mortality predicted risk (%)	0.86%	10.32%	25.22%	38.63%
Drug-related mortality predicted risk (%)	0.42%	3.90%	5.44%	24.26%

services in England; services with fewer deaths had wider variation in Harrel's C statistic than services with more events. Four services had no deaths over the studied timeframe and were not included in the meta-analysis. The summary-pooled Harrel's C statistic was 0.77 (95% CI 0.75–0.79), ranging from 0.66 (95% CI 0.57–0.74) to 0.99 (95% CI 0.98–1.00). The I^2 value (i.e. the percentage of total variation in Harrel's C statistics explained by between service heterogeneity) was 99.5% (95% CI 96.0–99.8%).

Drug-related mortality

The optimism-adjusted final model explained 21% of the variation in time to drug-related mortality (R^2_D), the D statistic was 1.07 and Harrell's C statistic was 0.74. Supplementary Fig. 3 shows a forest plot of Harrel's C statistics. Six services had no drug-related deaths over the studied timeframe and were not included in the meta-analysis. The summary-pooled Harrel's C statistic was 0.81 (95% CI 0.80–0.82), ranging from 0.63 (95% CI 0.47–0.79) to 0.99 (95% CI 0.99–1.00). The I^2 value was 99.9% (95% CI 99.7–100.0%).

Calibration

All-cause mortality

The optimism-adjusted calibration slope was 1.01 (0.95–1.08). The observed all-cause mortality risk at 6 months was 1.03% (0.99%–1.07%), compared with a mean individual predicted risk of 5.02% (5.00–5.04%).

Drug-related mortality

The optimism-adjusted calibration slope was 1.01 (0.94–1.10). The observed drug-related mortality risk at 6 months was 0.55% (0.52–0.58%), compared with a mean individual predicted risk of 2.19% (2.18–2.20%).

Table 5 depicts clinical examples of 6-month all-cause and drug-related mortality risk for individuals with OUD presenting to community addiction services in England. A web calculator for individualised risk is available at https://connect.calcapp.net/?app=4pekem.

Discussion

We have developed and internally validated Roberts' OUD mortality risk using two multivariable prognostic models to estimate 6-month all-cause and drug-related mortality risk for people with OUD presenting for baseline assessment at community addiction services in England. To our knowledge, no previous models have been developed to examine these outcomes in the studied population. Therefore, our models may provide clinically useful information and assistance to both patients and professionals when making treatment and care decisions.

Both models performed well in terms of discriminatory ability, with optimism-adjusted Harrell's C above 0.7 and D statistics above 1.0. Only around one-fifth of the variation in time to death was explained by each model. Although addition of further predictors has the potential to increase this statistic, we were mindful throughout model development and consultation with clinicians and patients of the need for balance between model parsimony and performance. Models that have excessive numbers of parameters or use parameters that may require complex questioning or interpretation are likely to face implementation problems and thus lack clinical utility within time-pressured services, particularly if risk scoring is completed by professionals with a range of levels of experience and training.[30] As there was substantial heterogeneity in discriminative ability among services across England, the pooled Harrel's C statistics should be interpreted with caution owing to the large number of services with a low EPV rate; the pooled estimate is likely to have been an overestimate of discriminative performance. However, it was reassuring that all individual service Harrel's C estimates remained >0.65 for both models. Patient and clinician consultation before model development suggested that a ten-variable model would be optimal. Both final models contain 11 variables or fewer and are thus roughly in accord with this consensus.[30] Mean predicted mortality risks were higher in both models than the observed risk; this is probably indicative of the risk reduction associated with being engaged in addiction treatment, in particular, for individuals receiving opiate agonist therapy (OAT). Indeed, recent estimates demonstrate a three and a half times higher drug-related death rate for those not in receipt of OAT compared with those on OAT after adjustment for confounders,[18] and a fully adjusted *post hoc* model, which we developed to assess the impact of the binary variable of whether individuals had remained engaged in addiction treatment at 6 months from baseline assessment, demonstrated a significantly decreased risk of both all-cause and drug-related mortality at 6 months if individuals remained in treatment. As these models are intended to calculate risk at baseline assessment in community addiction services – a point at which, by definition, individuals are currently not in receipt of any treatment, including OAT – they allow individualised risk calculation at a clinically important point in time and allow individuals and professionals to contemplate risks without subsequent treatment. Provision of individualised risk information at this juncture may thus promote internally generated behaviour change to address modifiable risk factors for individuals with OUD and increase the likelihood of their remaining engaged in

https://doi.org/10.1192/bjp.2025.10313 Published online by Cambridge University Press

treatment,[31] as well as assisting professionals in the prioritisation of finite resources such as prescriber availability and ensuring that individuals with elevated risks are actively supported to remain in treatment services.[32] There may be concern that providing individualised mortality risk could result in emotional distress for both patients and professionals; however, patient and clinician involvement suggested that this information would be welcomed, compared with the current clinical reality of stating that there is increased risk but with little personalised quantification. Indeed, studies in other settings have demonstrated perceived utility of provision of this type of individualised health information.[33] It is notable that no fractional polynomial of any degree resulted in the inclusion of age in the drug-related mortality model; this is in accordance with previous research that has challenged the 'ageing cohort' theory, indicating instead that the observed increase in drug-related deaths does not appear to be driven by age.[34]

This study had several strengths, including the comprehensive, non-selective and national nature of the data-set and the substantial involvement of clinicians and patients from the outset and throughout model development, validation and interpretation. The pre-publication of the protocol, alongside documentation of any deviation, and reporting of the study in accordance with the TRIPOD + AI statement also are notable strengths.[35] However, there were also several potential limitations. All prognostic indicator variables were collected retrospectively from an administrative data-set, the underlying data for which was supplied by addiction treatment, hospital and death registration services. There was therefore potential information bias and risk of lack of availability of some predictor variables, as noted particularly with HIV and Hepatitis C, if submitted documentation was incomplete. Although relying on routinely documented clinical information as the source of prognostic information has limitations, this approach has been used frequently and reflects how the model would be used in clinical practice; for instance, some information may not be available to professionals or patients at the time of baseline assessment.[36] The model will benefit from independent validation in other samples and subsequent examination of its utility in clinical practice and its acceptability among professional and patient groups; potentially suitable data-sets have been identified in both Wales and Australia.[37,38] Given this, we have ensured reporting of model baseline survivor and linear predictor values. Continued co-production through independent validation and implementation with clinicians and patients will remain a key requirement.

Despite significant expansion and understanding in the use of machine learning methods to develop prognostic models across healthcare sectors, initial patient and clinician feedback has demonstrated reticence with respect to employing these in the context of mortality prediction in OUD. The perception of a 'black box' or lack of transparent understanding of what prediction outcome scores are based on, as well as the relative infancy of clinical informatics within the OUD field, led to concerns about clinical utility and implementation within community addiction services.[39] Clinicians were comfortable with clinical risk tools developed using classical statistical methods and their corollaries used in other areas of healthcare[36] and welcomed their potential expansion within addiction settings. However, there was concern among patients that results from machine learning methods would not be believed and that explanation of algorithms could create difficulties in conveying the predictive information to individuals accessing services. As such, traditional statistical methods were chosen to develop this model.

Standardised all-cause and drug-related mortality rates are significantly elevated among people with OUD, and, despite a significant body of literature describing individual prognostic risk factors, often clinical judgement alone is used to consider individual prognosis and the prioritisation of treatment interventions in addiction treatment services. Whereas other areas of medicine routinely incorporate risk tools into care to assist clinical decision-making,[36,40] clinical informatics within the addiction field has been somewhat slower to progress. Given the significant elevated mortality risks within this population, the development and validation of individualised prediction models that demonstrate good optimism-adjusted discrimination and calibration is timely, warranted and urgent. This is the first stage of model assessment. Independent validation and demonstration of the clinical utility of both models are necessary next steps to ensure buy-in from professionals, policy makers and patients and for the models to be valued and successfully implemented.

Emmert Roberts, PhD, National Addiction Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; South London and the Maudsley NHS Foundation Trust, London, UK; and Department of Health and Social Care, Office for Health Improvement and Disparities, London, UK; **John Strang**, MD, South London and the Maudsley NHS Foundation Trust, London, UK; and National Addiction Centre, Institute of Psychiatry, Psychology and Neuroscience, Kings College London, London, UK; **Eve Taylor**, PhD, National Addiction Centre, Institute of Psychiatry, Psychology and Neuroscience, Kings College London, London, UK; **Jamie Crummy**, PhD, Department of Health and Social Care, Office for Health Improvement and Disparities, London, UK; **Tim Lowden**, MSc, Department of Health and Social Care, Office for Health Improvement and Disparities, London, UK; **Chioma Amasiatu**, MSc, Department of Health and Social Care, Office for Health Improvement and Disparities, London, UK; **Brian Eastwood**, PhD, Department of Health and Social Care, Office for Health Improvement and Disparities, London, UK

Correspondence: Emmert Roberts. Email: emmert.roberts@kcl.ac.uk

First received 2 Feb 2025, final revision 28 Apr 2025, accepted 17 May 2025

Supplementary material

The supplementary material is available online at https://doi.org/10.1192/bjp.2025.10313

Data availability

The data-sets generated and analysed during the current study are not publicly available as they contain sensitive patient-identifiable data. Although access to the linked data-set is only available within the DHSC subject to approval, aggregated extracts of NDTMS data are publicly available at www.ndtms.net, and extracts of HES-ONS data are available through the Data Access Request Service at NHS England. The full published protocol and statistical analysis plan are also available.[13]

Author contributions

All authors met all criteria for authorship according to the ICMJE recommendations. Contribution of article co-authors as per the Contributor Roles Taxonomy (CRediT) author statement were as follows. E.R.: conceptualisation, methodology, formal analysis, investigation, data curation, writing – original draft, writing – review and editing, visualisation, and project administration; J.S.: writing – review and editing, and supervision; E.T.: writing – review and editing; J.C.: methodology, project administration, and writing – review and editing; T.L.: methodology, project administration, and writing – review and editing; C.A.: methodology, project administration, and writing – review and editing; B.E.: conceptualisation, methodology, data curation, writing – review and editing, and supervision.

Declaration of interest

All authors have completed the ICJME Unified Competing Interest form (available on request from the corresponding author). E.R. is a member of the *British Journal of Psychiatry* editorial board; he did not take part in the review or decision-making process for this paper.

Funding

This paper represents independent research funded by the corresponding author's National Institute of Health and Care Research Advanced Fellowship (NIHR302215). The funders had no contribution to the study design; the collection, analysis and interpretation of data; the writing of the report; or the decision to submit the article for publication. All authors were independent from funders, had full access to all of the data (including statistical reports and tables) in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the DHSC.

https://doi.org/10.1192/bjp.2025.10313 Published online by Cambridge University Press

References

1 Office for National Statistics (ONS). *Deaths Related to Drug Poisoning in England and Wales: 2021 Registrations*. ONS, 2022 (https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsrelatedtodrugpoisoninginenglandandwales/2021registrations).

2 Office for Health Improvement & Disparities. *Adult Substance Misuse Treatment Statistics 2022 to 2023: Report*. TSO (The Stationery Office), 2023 (https://www.gov.uk/government/statistics/substance-misuse-treatment-for-adults-statistics-2022-to-2023/adult-substance-misuse-treatment-statistics-2022-to-2023-report).

3 Larney S, Tran LT, Leung J, Santo T, Santomauro D, Hickman M, et al. All-cause and cause-specific mortality among people using extramedical opioids: a systematic review and meta-analysis. *JAMA Psychiatry* 2020; **77**: 493–502.

4 Bahji A, Cheng B, Gray S, Stuart H. Mortality among people with opioid use disorder: a systematic review and meta-analysis. *J Addict Med* 2020; **14**: E118–32.

5 Castaneda C, Nalley K, Mannion C, Bhattacharyya P, Blake P, Pecora A, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. *J Clin Bioinf* 2015: **5**: 1–16.

6 Sordo L, Barrio G, Bravo MJ, Indave BI, Degenhardt L, Wiessing L, et al. Mortality risk during and after opioid substitution treatment: systematic review and meta-analysis of cohort studies. *BMJ* 2017: **357**: j1550.

7 Webster LR. Risk factors for opioid-use disorder and overdose. *Anesth Analges* 2017: **125**: 1741–8.

8 Lyons RM, Yule AM, Schiff D, Bagley SM, Wilens TE. Risk factors for drug overdose in young people: a systematic review of the literature. *J Child Adolesc Psychopharmacol* 2019; **29**: 487–97.

9 Tseregounis IE, Henry SG. Assessing opioid overdose risk: a review of clinical prediction models utilizing patient-level data. *Transl Res* 2021; **234**: 74–87.

10 Garbin C, Marques N, Marques O. Machine learning for predicting opioid use disorder from healthcare data: a systematic review. *Comput Methods Progr Biomed* 2023; **236**: 107573.

11 Roberts E, Doidge JC, Harron KL, Hotopf M, Knight J, White M, et al. National administrative record linkage between specialist community drug and alcohol treatment data (the National Drug Treatment Monitoring System (NDTMS)) and inpatient hospitalisation data (Hospital Episode Statistics (HES)) in England: design, method and evaluation. *BMJ Open* 2020; **10**: e043540.

12 NHS England. *Primary Medical Care Policy and Guidance Manual (PGM)*. NHS England, 2017 (https://www.england.nhs.uk/publication/primary-medical-care-policy-and-guidance-manual-pgm/).

13 Roberts E, Strang J, Horgan P, Eastwood B. The development and internal validation of a multivariable model predicting 6-month mortality for people with opioid use disorder presenting to community drug services in England: a protocol. *Diagn Progn Res* 2024; **8**: 7.

14 Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024; **385**: e078378.

15 Jewell A, Pritchard M, Barrett K, Green P, Markham S, McKenzie S, et al. The Maudsley Biomedical Research Centre (BRC) data linkage service user and carer advisory group: creating and sustaining a successful patient and public involvement group to guide research in a complex area. *Res Involv Engagem* 2019; **5**: 1–10.

16 NHS Digital. *National Drug Treatment Monitoring System*. NHS Digital, 2004 (https://digital.nhs.uk/data-and-information/information-standards/governance/latest-activity/standards-and-collections/dapb0107-national-drug-treatment-monitoring-system).

17 NHS Digital. *Hospital Admitted Patient Care Activity*. NHS Digital, 2024 (https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity).

18 McAuley A, Fraser R, Glancy M, Yeung A, Jones HE, Vickerman P, et al. Mortality among individuals prescribed opioid-agonist therapy in Scotland, UK, 2011-20: a national retrospective cohort study. *Lancet Public Health* 2023; **8**: e484–93.

19 NHS Digital. *Hospital Episode Statistics Data Dictionary*. NHS Digital, 2025 (https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary).

20 National Drug Treatment Monitoring System. *Events and Training Material*. Office for Health Improvement and Disparities and University of Manchester, 2024 (https://www.ndtms.net/Events).

21 Riley RD, Snell KIE, Ensor J, Burke DL, Harrell Jr FE, Moons KGM et al. Minimum sample size for developing a multivariable prediction model: PART II-binary and time-to-event outcomes. *Stat Med* 2019; **38**: 1276–96.

22 Ensor J, Martin EC, Riley RD. *Package 'pmsampsize'*. The Comprehensive R Archive Network, 2022.

23 White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011: **30**: 377–99.

24 Rubin DB, Schenker N. Multiple imputation in health-are databases: an overview and some applications. *Stat Med* 1991: **10**: 585–98.

25 Lin DY, Wei L-J. The robust inference for the Cox proportional hazards model. *J Am Stat Assoc* 1989; **84**: 1074–8.

26 Heinze G, Wallisch C, Dunkler D. Variable selection-a review and recommendations for the practicing statistician. *Biom J* 2018; **60**: 431–49.

27 Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009; **338**: b604.

28 Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* 2016; **353**: i3140.

29 Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; **15**: 361–87.

30 Markowetz F. All models are wrong and yours are useless: making clinical prediction models impactful for patients. *NPJ Precis Oncol* 2024; **8**: 54.

31 French DP, Cameron E, Benton JS, Deaton C, Harvie M. Can communicating personalised disease risk promote healthy behaviour change? A systematic review of systematic reviews. *Ann Behav Med* 2017; **51**: 718–29.

32 Tai AMY, Kim JJ, Schmeckenbecher J, Kitchin V, Wang J, Kazemi A, et al. Clinical decision support systems in addiction and concurrent disorders: a systematic review and meta-analysis. *J Eval Clin Pract* 2024; **30**: 1664–83.

33 Nathan R, Bhandari S. Risk assessment in clinical practice: a framework for decision-making in real-world complex systems. *BJPsych Adv* 2024; **30**: 53–63.

34 Lewer D, Brothers TD, Van Hest N, Hickman M, Holland A, Padmanathan P, et al. Causes of death among people who used illicit opioids in England, 2001–18: a matched cohort study. *Lancet Public Health* 2022; **7**: e126–35.

35 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. *Circulation* 2015; **131**: 211–9.

36 Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ* 2007; **335**: 136.

37 Larney S, Jones N, Fiellin DA, Nielsen S, Hickman M, Dobbins T, et al. Data resource profile: the opioid agonist treatment and safety (OATS) study, New South Wales. *Austr Int J Epidemiol* 2020; **49**: 1774–5.

38 Trefan L, Akbari A, Paranjothy S, Farewell DM, Gartner A, Fone D, et al. Electronic Longitudinal Alcohol Study in Communities (ELAStiC) Wales – protocol for platform development. *Int J Popul Data Sci* 2019; **4**: 581.

39 Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Mach Intell* 2019; **1**: 206–15.

40 Olesen JB, Torp-Pedersen C, Hansen ML, Lip GY. The value of the CHA2DS2-VASc score for refining stroke risk stratification in patients with atrial fibrillation with a CHADS2 score 0–1: a nationwide cohort study. *Thromb Haemost* 2012; **107**: 1172–9.

https://doi.org/10.1192/bjp.2025.10313 Published online by Cambridge University Press