Accepted manuscript: Authors' Copy

Self-**HA**rm one-year **R**isk **P**rediction in personality disorders (SHARP-PD): a clinical risk prediction model development and internal-external cross-validation study

Joe Kwun Nam Chan[1], Ivan Wai Lok Chu[1], Corine Sau Man Wong[2], Sai Yu Simon Lui[1,3], Pak Chung Sham[1], Hon Cheong So[4], Wing Chung Chang[1,2*]

[1]Department of Psychiatry, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong

[2] School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, China

[3] Department of Psychiatry, Queen Mary Hospital, Hong Kong, China

[4] School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong, China

*Corresponding authors**:** Wing Chung CHANG

Department of Psychiatry, The University of Hong Kong

Queen Mary Hospital, Pokfulam, Hong Kong

Tel: (852) 22554486 Fax: (852) 28551345

Email: changwc@hku.hk

Abstract

Background. Personality-disorders (PDs) are associated with markedly-increased risk of self-harm. We aimed to develop and validate a risk-prediction model (**S**elf-**HA**rm one-year **R**isk

This peer-reviewed article has been accepted for publication but not yet copyedited or typeset, and so may be subject to change during the production process. The article is considered published and may be cited using its DOI.

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

Prediction in personality disorders [SHARP-PD]) to predict 1-year self-harm in people with newly-diagnosed PD.

Methods. This population-based study identified individuals with first-recorded diagnosis of PD in 2002–2021, utilising an electronic-health-record database of public-healthcare service in Hong-Kong. SHARP-PD was developed using binary-logistic-regression model, with Least-Absolute-Shrinkage-and-Selection-Operator penalisation. Model performance was assessed through internal-external cross-validation (IECV), sequentially leaving out one cluster from the seven healthcare-catchment areas for validation and averaging performance across seven clusters via random-effect meta-analysis. Predictors for model development included socio-demographic characteristics, psychiatric and physical morbidities, psychotropic medication use, history of self-poisoning and self-inflicting injuries, and psychiatric service utilisation.

Results. Among 6229 patients with PD, 608 (9.8%) performed self-harm at 1-year post-diagnosis. The model exhibited excellent performance in IECV: area-under-the-curve 0.93 (0.92–0.94); Brier-score 0.06 (0.05–0.06); calibration-slope 1.01 (0.95–1.08); calibration-intercept 0.02 (-0.14–0.19); integrated-calibration-index 0.15 (0.11–0.19); observed/expected-ratio 1.00 (0.92–1.08). Decision-curve-analysis showed that SHARP-PD provided greater net-benefit than the treat-all and treat-none strategies starting from the threshold probability of 1%. The model performance remained largely-consistent in sensitivity-analyses excluding predictors with near-zero prevalence, complete-case analyses, including longer washout period for incident PD diagnosis, and restricting to non-fatal self-harm in outcome definition, and across key socio-demographic subgroups of sex, age categories, and borderline PD.

Conclusions. The novel SHARP-PD enables systematic early-identification of high-risk individuals, and can inform targeted allocation of resource-intensive, preventive interventions for self-harm in PD.

Keywords: Risk prediction model, personality disorders, self-harm, electronic health record, internal-external cross-validation

Introduction

Personality disorders (PDs) are prevalent mental disorders, affecting 7.8% of the general population [1]. PDs, particularly borderline PD, are strongly associated with suicide and self-injurious behaviours [2]. It is estimated that 65–80% of individuals with borderline PD engage in some forms of self-harm [2]. In a clinical cohort of 394 patients with borderline PD, 75% had attempted suicide and 90% had engaged in non-fatal self-injury [3]. A recent meta-analysis of 79 studies found that PD was associated with 4-fold increase in all-cause mortality and a 9-fold increase in suicide, compared to the general population [4]. These findings highlight PD as a major driver of self-harm and premature mortality, underscoring the need for earlier detection, regular and ongoing assessment of self-harm and suicide risk, and systematic implementation of evidence-based psychological and crisis intervention across both mental health and general medical services.

Recently, increasing attention has been given to developing risk-prediction models to estimate individualised risk of self-harm, with an aim to effectively identify high-risk patients and target preventive interventions. The OxMIS model, developed using data from Swedish national registers, predicts 1-year suicide risk in people with severe mental disorders and achieved an area-under-the-curve (AUC) value of 0.71 at external validation [5]. Its strongest predictors included inpatient status at diagnosis, previous self-harm, and male sex. The Simon-model, derived from US population-based electronic-health records (EHR), predicts 90-day self-harm following

psychiatric and general medical visits, with an external validation AUC of 0.86 [6,7]. Key predictors included depression, prior self-harm, and substance-use disorders. Notably, a recent review found that only small number of existing risk-prediction tools for self-harm have undergone external validation, and even fewer had demonstrated adequate discrimination and calibration in independent samples, limiting the generalisability of the models [8]. Many other models are also constrained by inadequate assessment of predictive performance, small sample size, suboptimal handling of missing data, and lack of effective methods to sufficiently address overfitting and optimism [8]. Importantly, none of existing tools has been specifically developed and validated for people with PD, despite their substantially-elevated risk of self-harm and distinctive patterns of service use. Therefore, current models offer limited guidance for clinical-decision making in this respect. Given the high clinical burden and frequent contact of people with PD with both mental health and general medical services, a scalable, well-validated, PD-specific prediction model for self-harm would meaningfully enhance risk stratification, guide allocation of interventions, and support more precise, preventive psychiatry in routine care.

To this end, we conducted this large, population-based cohort study to develop and validate a 1-year risk prediction tool for self-harm in individuals with newly-diagnosed PD in real-world care: **S**elf-**HA**rm one-year **R**isk **P**rediction in personality disorders (SHARP-PD). We utilised a territory-wide EHR database of public healthcare system, enabling comprehensive capture of self-harm events, and routinely-collected characteristics in prediction. Internal-external cross-validation (IECV), sequentially leaving out each of the seven healthcare catchment areas as a validation dataset and averaging performance across areas, was applied to provide a rigorous assessment of model transportability. Sensitivity analyses were performed to test robustness of the model, including excluding predictors with near-zero prevalence, conducting complete-case

analyses, and applying a longer washout period for PD diagnosis. We also examined model performance in key sociodemographic subgroups (females vs males, age <25 vs ≥25 years and borderline PD) to assess its applicability across clinically relevant groups.

Methods

Study design and data source

Patient data were retrieved from the Clinical-Data-Analysis-and-Reporting-System (CDARS) [9], a territory-wide EHR database developed by the Hospital Authority (HA), which is a statutory body delivering government-subsidised, universal health coverage to all HK residents (approximately 92% being Chinese) by managing all public hospitals, specialists and general outpatient clinics in HK. A detailed description of CDARS has been reported elsewhere [10]. Briefly, CDARS is an integrated, longitudinal patient electronic-record system capturing clinical data across all healthcare settings of HA facilities since 1995. The database contains information on patients' demographic and clinical characteristics comprising details of diagnoses, visits at emergency departments and outpatient clinics, hospital admissions, and laboratory results. It also contains data on deaths through its internal linkage to regional death registry from the Immigration Department. Clinical Data are collected and entered into the computerised clinical management system by treating clinicians and other healthcare professionals, and are transferred to CDARS for audit and research purposes. CDARS generates unique, anonymised patient identifiers to protect patient privacy and to link all medical records. This database has been used successfully to conduct high-quality research on psychiatric disorders, and development of risk-prediction models [11-15].

Study population and patient identification

The study period was between 1-January-2002 and 31-December-2021. We identified all individuals aged ≥12 years with a first-recorded PD diagnosis for public psychiatric inpatient

admissions or outpatient care in HA facilities as the study population. Diagnosis of PD was ascertained by the International Classification of Disease, 10th revision (ICD10: F60). Follow-up of patients began on the diagnosis of PD (i.e., first-recorded PD discharge diagnosis from psychiatric admission or first-recorded PD diagnosis in specialist psychiatric outpatient clinic, whichever came earlier), and lasted until one year of follow-up, occurrence of self-harm, leaving the HK health service, death or 31-December-2021, whichever came first. To ensure inclusion of only incident PD cases, individuals with any recorded diagnosis of PD prior to 1-January-2002 (computerised service for psychiatric services implemented since January-1-2000) in HA psychiatric inpatient or outpatient records were excluded from the study population. This study was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (certification number: UW 23-540). Since this study was conducted on data from an EHR database with strictly-anonymised patient records, the requirement for informed consent was waived.

Outcomes and predictors

The study outcome was non-fatal self-harm and/or completed suicide within one year after first-recorded PD diagnosis. During the data extraction process, we ensured that the date of first-recorded self-harm outcome in the follow-up period is required to be later than the index diagnosis date (i.e., the difference between these dates is strictly greater than zero), and self-harm that occurred with the same psychiatric admission/ emergency department stay for first-recorded PD diagnosis was not classified as an outcome. Given the difficulties in determining the intent of self-harm events, we followed the approach of previous EHR-based research and adopted a broader definition of self-harm [13,14], including all self-poisoning and self-injurious behaviours with and without suicidal intent, and that were intentionally and unintentionally self-inflicted. ICD10 and

ICD9-CM codes of self-harm were used to identify non-fatal self-harm (ICD10: X60-84, Y10-Y34, Y87.0, Y87.2, T39, T42-43; ICD9 CM: E950-952, E980-E982, E953-E959, E983-E989). ICD10 codes of suicide were used to identify completed suicide (X60–X84, Y87.0).

Taking into consideration the risk factors identified in literatures on self-harm prediction, and the availability of clinical information that could be adequately captured in the EHR database, we selected an array of candidate predictors measuring information recorded on or before the prediction landmark (i.e., date of first-recorded PD diagnosis), including socio-demographic characteristics, diagnosis of borderline PD subtype, psychiatric and physical morbidities, psychotropic medication use, history of self-poisoning and self-inflicting injuries, and psychiatric service utilisation. Socio-demographics comprised age at PD diagnosis, sex, education level, and receipt of social/disability allowance (as a proxy for socioeconomic status). Psychiatric morbidities encompassed alcohol and substance-use disorders, schizophrenia-spectrum disorders, depressive disorder, bipolar disorder, anxiety disorders, obsessive-compulsive disorders, stress-related disorders, somatoform and conversion disorders, eating disorders, autism-spectrum disorders, and attention-deficit/hyperactivity disorder. Physical comorbidities were assessed using Charlson-Comorbidity Index, a widely-used measure of chronic physical comorbidity burden based on 17 physical conditions [16]. Psychotropic medications entailed first- and second-generation antipsychotics, antidepressants, lithium, mood-stabilising anticonvulsants, benzodiazepines, and Z-drugs. Psychiatric service utilisation within one year before the PD diagnosis was captured, including psychiatric hospitalisation, specialist-psychiatric-outpatient visits, clinical-psychology outpatient visits, and community psychiatric service contacts. Diagnoses of physical and psychiatric morbidities were ascertained using ICD codes (Supplementary Table 1).

Statistical analysis

Continuous and categorical variables for the overall sample and for each healthcare service catchment area (i.e., cluster) were compared using ANOVA and χ^2 tests, as appropriate. Characterising between-cluster heterogeneity is important as greater diversity across clusters enhances the assessment of how well model performance might generalise to other settings. Model development and validation followed contemporary guidelines of Collins et al. [17], and Riley et al. [18].

All pre-processing was embedded within the cross-validation framework to avoid data leakage. Single imputation with random forest was applied to the only variable within missing data (i.e., education level). SHARP-PD was developed using a binary logistic-regression model with Least-Absolute-Shrinkage-and-Selection-Operator (LASSO) penalisation. LASSO performed predictor selection and coefficient shrinkage, with the penalty parameter (λ) chosen to minimise the mean cross-validation error via k-fold cross-validation (i.e., λ and predictor selection determined in the development data and evaluated in the validation data). Predictors with coefficients shrunk to zero were excluded from the final model. Internal-external cross-validation (IECV) was applied to evaluate model performance and transportability. The dataset was partitioned into seven clusters based on healthcare-service catchment area (Hong Kong East, Hong Kong West, Kowloon East, Kowloon Central, Kowloon West, New Territories East, New Territories West). In each iteration, the model was trained on six clusters (development data) and validated on the remaining cluster (validation data), yielding seven sets of performance and calibration metrices. Compared with a simple random split of the data, IECV provides a more rigorous assessment model transportability as each validation fold represents a distinct healthcare cluster area, approximating multiple external validations, thereby reducing optimism in performance metrices.

Performance metrices across the seven internal-external validation folds were pooled using inverse-variance random-effects meta-analysis to obtain mean estimates and 95% confidence intervals (CIs), accounting for differences in sample size and potentially heterogeneity between clusters. After validation, the model was refitted on the overall sample using the median λ value selected across IECV folds to derive final coefficients for future external validation and clinical implementation. Coefficients were also transformed to odds ratios (ORs) for easer interpretation. Model performance was assessed using the AUC, where values of 0.5 and 1.0 indicate random performance and perfect discrimination, respectively. An AUC>0.7 was considered acceptable. Overall model fit was evaluated using the Brier score, defined as the mean squared difference between predicted and observed probabilities of 1-year self-harm, with lower values indicating better performance. Calibration, which measured the agreement between predicted and observed risk, was assessed using the calibration slope (target value=1), calibration intercept (i.e., calibration-in-the-large; target value=0), the integrated-calibration index (ICI; lower values indicate better calibration), and the observed/expected (O/E) risk ratio (target value=1). A calibration slope <1 suggests overfitting with over-extreme predictions, whereas a slope >1 suggests underfitting with under-extreme predictions. A negative calibration intercept indicates overestimation of risk on average, and a positive intercept indicates underestimation. Higher ICI values indicate greater average discrepancies between predicted and observed risks across the spectrum of predicted probabilities. An O/E ratio <1 indicates that the model overestimates risk on average, whereas an O/E ratio >1 indicates underestimation.

Potential clinical utility was assessed using decision-curve analysis within the IECV framework, producing seven net-benefit curves that were subsequently pooled to derive mean decision curves. Decision-curve analysis (DCA) quantifies the clinical net-benefit of using SHARP-PD across a

range of risk thresholds as a weighted balance between correctly identifying individuals who will self-harm (true positives) and incorrectly labelling and intervening on individuals who will not (false positives). The DCA is based on the assumption that there is an available intervention for self-harm that confers some benefit (e.g., dialectical behaviour therapy and related evidence-based psychosocial interventions for individuals with PD and self-harm), and that this benefit outweighs the potential harms and costs of unnecessary intervention, to justify the use of DCA in this clinical context. Net benefit was plotted across threshold probabilities from 0% to 50% for 1-year self-harm, as clinicians are unlikely to require a prediction model when their intervention threshold for self-harm is so high that exceeds 50%. The net benefit of SHARP-PD was compared with the default strategies of "treat all" and "treat none" across this clinically relevant range of thresholds. The standardized net benefit was calculated as the net benefit divided by the outcome prevalence, as the proportion of the maximum achievable clinical benefit that a hypothetical perfect model (which correctly identifies all cases and never treat non-cases) could deliver at a given threshold. The receiver operating characteristic curve was performed with the Youden index indicating the optimal predicted probability threshold for classifying the high-risk individuals.

Four sensitivity analyses were conducted. First, to reduce model instability, predictors with near-zero prevalence in the overall dataset were excluded and the model was refitted. Second, to evaluate the impact of imputation, the model was rerun in the complete-case dataset. Third, to strengthen the definition of incident PD, we extended the washout period for prior PD diagnoses to in or before 2005, and reran the model in individuals with first-recorded PD diagnoses in 2006–2021. Fourth, to enhance the outcome homogeneity (given only a minority of self-harm cases were complete suicide [n=51; 8.4% of self-harm outcomes], we restricted the outcome definition to non-fatal self-harm only. In addition, subgroup analyses were performed to further internally-externally

validate SHARP-PD within key strata: (1) females and males, (2) age<25 years and ≥25 years, and (3) borderline PD. All analyses were conducted in R (version 4.5.1), and two-sided P-values <0.05 were considered statistically significant.

Risk of bias and applicability

We assessed the risk of bias and applicability of SHARP-PD using the PROBAST tool [19], which was developed to evaluate risk of bias and applicability of diagnostic and prognostic prediction models in domains of participants, predictors, outcome and analysis (Supplementary Table 2).

Reporting

Reporting of the current study on the prediction model for 1-year self-harm in people with PD followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD+AI; Supplementary Table 3) [20].

Results

Sociodemographic and clinical characteristics

The study cohort comprised 6229 individuals diagnosed with incident PD (Table 1). Among the overall sample, the mean age at diagnosis was 35.7 years (SD=14.1), 63.2% female, 14.4% attained tertiary education or above, and 26.7% received social/disability allowance. A total of 56.7% of the sample was diagnosed with borderline PD subtype. The most common comorbid mental disorder was depressive disorder (30.8%), followed by stress-related disorders (21.5%) and substance-use disorders (16.3%). Antidepressants were the most frequently-prescribed psychotropic medications (65.0%), followed by benzodiazepines (54.8%) and Z-drugs (45.7%). One-fifth (19.2%) of individuals with PD had a history of self-poisoning, and 4.0% had a history of self-inflicting injuries. Differences across clusters for sociodemographic and clinical

characteristics are presented in Table 1. Missing value was only observed for education level (14.4%).

Development of SHARP-PD

Within the 1-year prediction horizon, the total number of self-harm events in the overall sample was 608 (9.8%; mean follow-up duration=3.9 months [SD=3.6]; Supplementary Table 4). The event-per-variable ratio was 19.6. From the 31 pre-specified candidate variables, the refitted LASSO logistic-regression in the overall sample using the median λ value selected across IECV folds retained 21 predictors, including male sex, tertiary education or above, diagnosis of borderline PD subtype, eating disorders, substance-use disorder, eating disorders, stress-related disorders, somatoform and conversion disorders, autism-spectrum disorders, depressive disorders, attention-deficit/hyperactivity disorder, history of self-inflicting injuries and self-poisoning, first-generation antipsychotics, antidepressants, Z-drugs, benzodiazepines, second-generation antipsychotics, lithium, clinical-psychology outpatient service visits, community psychiatric nurse contacts and psychiatric hospitalisation (Figure 1). Details of the risk equation for calculating individualised 1-year self-harm probabilities are provided in Supplementary Table 5. An interactive version of the model can be accessed online (https://sharpseries.shinyapps.io/pd_risk_app/).

Model performance

In IECV (Figure 2; Supplementary Table 6), SHARP-PD exhibited excellent model performance (AUC=0.93 [0.92–0.94]; Brier score=0.06 [0.05–0.06]) and calibration (calibration slope=1.01 [0.95–1.08]; calibration intercept=0.02 [-0.14–0.19]; ICI=0.15 [0.11–0.19]; O/E ratio=1.00 [0.92–1.08]). Graphically, the calibration curve aligned closely with the ideal calibration line (Figure 3). The model performance remained largely-consistent in sensitivity analyses excluding variables

with near-zero prevalence (AUC=0.93; Brier score=0.06; calibration slope=1.00; calibration intercept=0.02; ICI=0.15; OE ratio=1.01), complete-case analysis (AUC=0.93; Brier score=0.06; calibration slope=1.01; calibration intercept=0.02; ICI=0.15; OE ratio=1.01), with longer washout period (AUC=0.94; Brier score=0.06; calibration slope=1.02; calibration intercept=0.03; ICI=0.16; OE ratio=1.01), and restricting to non-fatal self-harm outcome only (AUC=0.94; Brier score=0.06; calibration slope=1.01; calibration intercept=0.03; ICI=0.15; OE ratio=1.01; Supplementary Table 7). No significant differences were observed in model performance and calibration of SHARP-PD for 1-year self-harm when stratified by sex and age categories, and in borderline PD.

Clinical usefulness

DCA demonstrated that SHARP-PD provided greater net benefit than the treat-all and treat-none strategies starting from the threshold probability of 1% (Figure 3). For example, if an intervention was considered for patients with predicted risk higher than 10%, the model provided net benefits of 0.076 (i.e., correctly identifying about 7.6 additional self-harm cases per 100 patients), corresponding to a standardized net benefit of 0.777 (i.e., achieving 77.7% of the maximum clinical benefit that a perfect model could deliver; Supplementary Table 8). Sensitivity and specificity values at pre-specified percentile of predicted probabilities are also reported in Supplementary Table 9. For example, at a top 15% of predicted risk (i.e., predicted risk=28.1%), the sensitivity and specificity for self-harm were 0.75 (0.72–0.79), and 0.92 (0.91–0.92), respectively. The ROC curve analysis indicated that 6.8% was the Youden-optimal predicted probability threshold for classifying high-risk individuals (Supplementary Figure 1). The distribution of the predicted probabilities of the overall PD sample was shown in Supplementary Figure 2.

Discussion

To our knowledge, this is the first study to develop and rigorously validate a clinical-risk prediction model for self-harm (including non-fatal self-harm and completed suicide) in people with newly-diagnosed PD. SHARP-PD showed excellent performance, with an AUC of 0.93, addressing an important unmet need in accurately identifying individuals with PD at highest risk of self-harm. The model was derived from a large, territory-wide, real-world clinical cohort within the public-healthcare system, supporting its potential generalisability at scale. Importantly, SHARP-PD relied solely on routinely-collected clinical characteristics and does not require additional investigations, bespoke assessments, or clinician-rated scales. This makes it well suited for integration into EHR systems, where relevant predictors can be extracted automatically and an individualised risk score generated at predefined clinical timepoints. Clinicians could receive automated alerts when a patient's predicted risk exceeds a prespecified threshold, facilitating earlier, targeted referral of high-risk individuals for more detailed assessment or resource-intensive preventive interventions. This clinical-prediction model might be conceived as an initial screening tool in the detection pathway capable of large-scale detection at low financial and labour costs, which may be particularly promising for middle-resource settings, where EHR infrastructures are expanded, not only for high-resource settings.

Using IECV, we validated SHARP-PD across seven distinct healthcare service clusters, with each cluster in turn serving as an independent validation sample. This approach mimics external validation by evaluating model performance in data from settings not used for model development. Although IECV is not equivalent to external validation in a completely independent health system, the results that SHARP-PD demonstrated consistently-excellent discrimination and calibration across all clusters provided evidence for the model's transportability and its feasibility for

implementation in HK or similar healthcare system. Similar IECV strategies have also been recently adopted in a transdiagnostic prediction model for psychotic and bipolar disorders by partitioning data across the five boroughs served by the South London and Maudsley (SLaM) National Health Service (NHS) Trust to obtain optimism-adjusted performance estimates [21]. We also conducted multiple sensitivity analyses to test the robustness of SHARP-PD, including excluding predictors with very low prevalence, performing complete-case analyses, and extending the washout period to address potential data noise during the early implementation of computerised psychiatric records. Across these analyses, the model retained excellent and comparable discrimination and calibration, further supporting its reliability for identifying high-risk individuals with PD in routine clinical practice.

To further minimise overfitting (i.e., the clinical-prediction model not generalising to populations outside of the development data), we applied the LASSO penalisation to shrink regression coefficients and perform variable selection. The tuning parameter was chosen according to the value yielding the minimum mean cross-validation error, thereby balancing model complexity and predictive performance. Regarding the variables finally retained in the model, history of self-harm emerged as the strongest predictor, consistent with prior work showing that suicidal behaviours in PD was best predicted by the number of previous attempts [22]. Our findings also suggested that, among methods of self-harm, self-poisoning conferred a substantially-higher subsequent risk compared to self-inflicted injuries, with an OR exceeding 30-fold. It is plausible that self-poisoning may reflect greater intent to die, higher medical lethality, or more pervasive access to means than less severe or more superficial self-injuries. This distinction has implications for clinical risk formulation in PD, emphasising the importance of method and lethality of previous self-harm acts rather than their mere occurrence.

We also observed that a broad range of comorbid mental disorders was associated with increased self-harm risk in PD. This extends previous study based on a primary-care sample highlighting comorbid alcohol misuse as a particular concern [23]. Several psychotropic medications were associated with increased self-harm. This partly accords with a recent Swedish study reporting elevated risk of suicidal behaviours with benzodiazepines, antidepressants, and antipsychotics in people with borderline PD [24]. These associations might be driven by indication and severity rather than direct causal effects. For instance, benzodiazepines are frequently prescribed to individuals with prominent affective instability, impulsivity or aggression [25,26], which are features closely linked to self-harm. Similarly, prescriptions of antidepressants, antipsychotics, and Z-drugs are often initiated or intensified during periods of acute crisis, emerging psychotic symptoms, insomnia, or severe mood disturbances, which themselves mark episodes of heightened risk. By contrast, the use of first-generation antipsychotics was associated with a reduced risk of self-harm. This counterintuitive finding may reflect differences in clinical indication (e.g., predominant psychotic symptoms rather than affective instability), or chance. It highlights the need for further pharmaco-epidemiological work examining antipsychotic treatment patterns and self-harm risk specifically in PD. Indicators of service use, including psychiatric hospitalisation, community psychiatric service contacts, and clinical-psychology outpatient visits in the year before PD diagnosis, were also positively associated with self-harm. These variables may capture greater symptom severity, functioning impairment and help-seeking, and may signal individuals with more complex and unstable clinical trajectories. The LASSO model also retained female sex and tertiary education or above as predictors associated with increased self-harm risk, despite a small degree of magnitude.

Notably, the presence and adverse impact of PD are frequently overlooked in clinical practice, where the diagnosis is often treated as one of exclusion rather than a condition warranting proactive, structured assessment and management [27]. This under-recognition extends to medical research [28] and policy, including the Global Burden of Disease [29], where the significance of PD is insufficiently acknowledged. Research funding for PD also remains disproportionately low, compared to other severe mental disorders [30]. These gaps in clinical awareness, research prioritisation, and funding collectively constrain the capacity of health systems to accurately identify, understand, and manage PD. This neglect is particularly concerning in light of a recent meta-analysis showing that the excess suicide-related mortality associated with PD, relative to the general population, appeared to be widening over time [4]. Against this backdrop, our study provides an important step towards precision psychiatry in PD, by generating an EHR-based, scalable clinical-prediction tool that stratifies 1-year self-harm risk at newly-diagnosed PD. By using routinely-collected data to generate individualised risk estimates, SHARP-PD can support more accurate identification of those at high risk of self-harm, guide the efficient allocation of resource-intensive interventions (e.g., specialist psychotherapy, crisis planning, assertive follow-ups), and embed preventive suicide strategies into routine mental health and general medical services. Specifically, the cut-off of predicted probability of ≥6.9%, as illustrated in the Youden index of the ROC curve analysis, may provide a reference threshold warranted for intervention. The current lack of PD-specific self-harm risk-screening tools underscores urgency of such approaches and highlights the potential clinical and public health value of implementing and further validating SHARP-PD in diverse healthcare systems.

Several limitations should be noted in this study. First, as in other EHR-based research, CDARS did not adequately capture the severity of psychiatric symptoms, which, however, may be a

relevant factor in risk prediction, thereby precluding us from incorporating clinical illness severity into our analysis. Second, medication exposure for psychotropic medications was determined based on dispensing records from CDARS, and thus patients' adherence to the prescribed regimens could not be guaranteed. Third, patient data were retrieved from medical-record database of public healthcare services managed by HA, patients who were under private psychiatric care were not included in the study. However, as HA is the predominant provider of psychiatric services to individuals with mental disorders in HK, the risk of selection bias or missing treated cases of PD should be minimal. Fourth, further implementation studies should be conducted and ethical challenges of communicating risk in real-world clinical practice must be carefully considered before clinical implementation of SHARP-PD.

In conclusion, this population-based cohort study developed and internally-externally validated the SHARP-PD model to stratify 1-year risk of self-harm in people with newly-diagnosed PD using territory-wide EHR data. SHARP-PD exhibited excellent predictive performance, with an AUC of 0.93 and excellent calibration. Its performance remained consistently high across all seven healthcare service clusters in IECV, providing strong evidence for the model's transportability and supporting its potential generalisability within similar secondary care systems. Routinely-collected demographic and clinical characteristics, including history and methods of self-harm, psychiatric comorbidities, psychotropic medication use, and recent psychiatric service utilisation, emerged as key predictors of self-harm risk in people with PD. Overall, SHARP-PD represents a novel, scalable, EHR-ready, risk-prediction model that can facilitate proactive, data-driven identification of high-risk individuals with PD for self-harm and more efficient allocation of preventive and therapeutic interventions in real-world clinical care.

Role of the funding source

This study was partially supported by Health and Medical Research Fund (21222111) and partially supported by a fellowship award from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU JRFS2526-7S04).

Conflict of Interest

The authors declare no competing interests. The authors alone are responsible for the content and writing of the paper.

Data availability

Data collected for this study are proprietary of the Hospital Authority of Hong Kong, which granted researchers permission and access to data. The data that support the findings of this study are available from this authority, but restrictions apply to the availability to these data (information on the cost is available from the corresponding author). The analytic codes are available from the corresponding author on request.

Acknowledgements

The authors would like to thank the colleagues in the Hospital Authority of Hong Kong for their kind assistance in data extraction for the current investigation. WC Chang and JKN Chan had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Figure 1: Predictors retained for self-harm at 1 year after diagnosis in the final LASSO-based logistic regression model. PD=personality disorders

Figure 2: Model performance for detecting self-harm at 1 year after diagnosis in people with personality disorders in real-world internal-external cross-validation. AUC=area under the receiver operating characteristic curve; CI=confidence intervals; REML=andom-effect meta-analysis.

Figure 3: Calibration plot (left) and decision-curve analysis (right) of the risk prediction model for self-harm at 1 year after diagnosis in people with personality disorders.

References

1. Winsper C, Bilgin A, Thompson A, Marwaha S, Chanen AM, Singh SP, Wang A, Furtado V. The prevalence of personality disorders in the community: a global systematic review and meta-analysis. Br J Psychiatry. 2020;216(2):69-78.
2. McClelland H, Cleare S, O'Connor RC. Suicide risk in personality disorders: a systematic review. Curr Psychiatry Rep. 2023;25:405-17.
3. Goodman M, Tomas IA, Temes CM, Fitzmaurice GM, Aguirre BA, Zanarini MC. Suicide attempts and self-injurious behaviours in adolescent and adult patients with borderline personality disorder. Personal Ment Health. 2017;11(3):157-163.
4. Chang WC, Chan JKN, Chu IWL, Yip APF, Chiu SSW, Yeung JWF, Wong CSM, Solmi M, Correll CU. All-cause and cause-specific mortality in people with personality disorders: a large-scale systematic review and meta-analysis of relative risk and aggravating or attenuating factors. World Psychiatry 2026 (Accepted).
5. Fazel S, Wolf A, Larsson H, Mallett S, Fanshawe TR. The prediction of suicide in severe mental illness: development and validation of a clinical prediction rule (OxMIS). Transl Psychiatry. 2019;9(1):98.

6. Simon GE, Johnson E, Lawrence JM, Rossom RC, Ahmedani B, Lynch FL, Beck A, Waitzfelder B, Ziebell R, Penfold RB, Shortreed SM. Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. Am J Psychiatry. 2018;175(10):951-960.
7. Walker RL, Shortreed SM, Ziebell RA, Johnson E, Boggs JM, Lynch FL, Daida YG, Ahmedani BK, Rossom R, Coleman KJ, Simon GE. Evaluation of electronic health record-based suicide risk prediction models on contemporary data. Appl Clin Inform. 2021;12(4):778-787.
8. Seyedsalehi A, Bailey J, Ogonah MGT, Fanshawe TR, Fazel S. Prediction models for self-harm and suicide: a systematic review and critical appraisal. BMC Med. 2025;23(1):549.
9. Hospital Authority Head Office IT Department. Clinical Data Analysis & Reporting System (CDARS) User's Manual: 2.0 ed. Hong Kong: Hong Kong Hospital Authority;2003:3.
10. Cheung NT, Fung V, Wong WN, Tong A, Sek A, Greyling A, Tse N, Fung H. Principles-based medical informatics for success--how Hong Kong built one of the world's largest integrated longitudinal electronic patient records. Stud Health Technol Inform. 2007;129(Pt 1):307-10.
11. Chan JKN, Solmi M, Correll CU, Wong CSM, Lo HKY, Lai FTT, Chang WC. Predicting 10-year risk of chronic kidney disease in lithium-treated patients with bipolar disorder: A risk model development and internal cross-validation study. Eur Neuropsychopharmacol. 2025;95:24-30.
12. Chan JKN, Solmi M, Correll CU, Wong CSM, Lo HKY, Lai FTT, Chang WC. Lithium for bipolar disorder and risk of thyroid dysfunction and chronic kidney disease. JAMA Netw Open. 2025;8(2):e2458608.

13. Chai Y, Luo H, Wong GHY, Tang JYM, Lam TC, Wong ICK, Yip PSF. Risk of self-harm after the diagnosis of psychiatric disorders in Hong Kong, 2000-10: a nested case-control study. Lancet Psychiatry. 2020;7(2):135-147.
14. Chai Y, Luo H, Yip PSF. Prevalence and risk factors for repetition of non-fatal self-harm in Hong Kong, 2002-2016: a population-based cohort study. Lancet Reg Health West Pac. 2020;2:100027.
15. Dong W, Wan EYF, Fong DYT, Tan KC, Tsui WW, Hui EM, Chan KH, Fung CSC, Lam CLK. Development and validation of 10-year risk prediction models of cardiovascular disease in Chinese type 2 diabetes mellitus patients in primary care using interpretable machine learning-based methods. Diabetes Obes Metab. 2024;26(9):3969-3987.
16. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol. 1992;45(6):613-619.
17. Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, Harrell FE Jr, Martin GP, Moons KGM, van Smeden M, Sperrin M, Bullock GS, Riley RD. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ. 2024;384:e074819.
18. Riley RD, Archer L, Snell KIE, Ensor J, Dhiman P, Martin GP, Bonnett LJ, Collins GS. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. BMJ. 2024;384:e074820.
19. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S; PROBAST Group†. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019 ;170(1):51-58.

20. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
21. Arribas M, de Micheli A, Krakowski K, Stahl D, Correll CU, Young AH, Andreassen OA, Vieta E, Arango C, McGuire P, Oliver D, Fusar-Poli P. Joint detection of risk for psychotic disorders or bipolar disorders in clinical practice in the UK: development and validation of a clinical prediction model. Lancet Psychiatry. 2026;13(1):14-23.
22. Soloff PH, Lis JA, Kelly T, Cornelius J, Ulrich R. Risk factors for suicidal behavior in borderline personality disorder. Am J Psychiatry. 1994;151(9):1316-23.
23. Doyle M, While D, Mok PL, Windfuhr K, Ashcroft DM, Kontopantelis E, Chew-Graham CA, Appleby L, Shaw J, Webb RT. Suicide risk in primary care patients diagnosed with a personality disorder: a nested case control study. BMC Fam Pract. 2016;17:106.
24. Lieslehto J, Tiihonen J, Lähteenvuo M, Mittendorfer-Rutz E, Tanskanen A, Taipale H. Comparative effectiveness of pharmacotherapies for the risk of attempted or completed suicide among persons with borderline personality disorder. JAMA Netw Open. 2023;6(6):e2317130.
25. Gardner DL, Cowdry RW. Alprazolam-induced dyscontrol in borderline personality disorder. Am J Psychiatry. 1985;142(1):98-100.
26. Cowdry RW, Gardner DL. Pharmacotherapy of borderline personality disorder. Alprazolam, carbamazepine, trifluoperazine, and tranylcypromine. Arch Gen Psychiatry. 1988;45(2):111-9.
27. Snowden P, Kane E. Personality disorder: no longer a diagnosis of exclusion. BJPsych Bull. 2003;27:401-3.

28. Tyrer P, Reed GM, Crawford MJ. Classification, assessment, prevalence, and effect of personality disorder. Lancet. 2015;385:717-26.

29. GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet. 2024;403:2133-61.

30. Zimmerman M, Gazarian D. Is research on borderline personality disorder underfunded by the National Institute of Health? Psychiatry Res. 2014;220(3):941-4.

Table 1. Baseline characteristics of the study sample

Characteristics	Overall (n = 6229)	HKE (n=423)	HKW (n=344)	KE (n=794)	KC (n=865)	KW (n=1131
Socio-demographics						
Age at PD diagnosis, mean (SD) year	35.7 (14.1)	37.3 (13.5)	36.8 (14.5)	36.6 (14.7)	37.2 (14.4)	36.5 (13
Sex, n (%)						
Male	2295 (36.8)	133 (31.4)	137 (39.8)	248 (31.2)	364 (42.1)	490 (43.
Female	3934 (63.2)	290 (68.6)	207 (60.2)	546 (68.8)	501 (57.9)	641 (56.
Education level, n (%)						
Secondary education or below	4804 (85.6)	255 (77.5)	230 (77.4)	625 (85.5)	672 (86.2)	908 (87.
Tertiary education or above	810 (14.4)	74 (22.5)	67 (22.6)	106 (14.5)	108 (13.8)	126 (12.
Receipt of social/disability allowance, n (%)	1663 (26.7)	99 (23.4)	72 (20.9)	235 (29.6)	265 (30.6)	361 (31.
Diagnosis of borderline PD subtype, n (%)	3529 (56.7)	246 (58.2)	160 (46.5)	488 (61.5)	439 (50.8)	575 (50.
Psychiatric comorbidities, n (%)						
Alcohol use disorder	482 (7.7)	48 (11.3)	31 (9.0)	61 (7.7)	54 (6.2)	89 (7.9)
Substance use disorders	1017 (16.3)	81 (19.1)	50 (14.5)	95 (12.0)	169 (19.5)	204 (18.
Schizophrenia-spectrum disorders	704 (11.3)	47 (11.1)	41 (11.9)	85 (10.7)	113 (13.1)	127 (11.
Depressive disorder	1918 (30.8)	147 (34.8)	119 (34.6)	251 (31.6)	247 (28.6)	304 (26.
Bipolar disorder	249 (4.0)	20 (4.7)	14 (4.1)	24 (3.0)	27 (3.1)	37 (3.3)
Anxiety disorders	256 (4.1)	23 (5.4)	25 (7.3)	32 (4.0)	40 (4.6)	32 (2.8)
Obsessive-compulsive disorders	99 (1.6)	14 (3.3)	6 (1.7)	11 (1.4)	16 (1.8)	15 (1.3)
Stress-related disorders	1339 (21.5)	100 (23.6)	71 (20.6)	212 (26.7)	186 (21.5)	228 (20.
Somatoform and conversion disorders	48 (0.8)	3 (0.7)	2 (0.6)	4 (0.5)	10 (1.2)	4 (0.4)
Eating disorders	78 (1.3)	9 (2.1)	0 (0.0)	11 (1.4)	9 (1.0)	15 (1.3)
Autism-spectrum disorders	43 (0.7)	2 (0.5)	4 (1.2)	3 (0.4)	7 (0.8)	7 (0.6)
Attention-deficit/hyperactivity disorder	67 (1.1)	4 (0.9)	5 (1.5)	4 (0.5)	8 (0.9)	13 (1.1)
Psychotropic medications, n (%)						
First-generation antipsychotics	2128 (34.2)	153 (36.2)	116 (33.7)	279 (35.1)	330 (38.2)	438 (38.
Second-generation antipsychotics	1752 (28.1)	150 (35.5)	106 (30.8)	190 (23.9)	241 (27.9)	323 (28.
Antidepressants	4047 (65.0)	303 (71.6)	217 (63.1)	536 (67.5)	529 (61.2)	717 (63.

Lithium	193 (3.1)	15 (3.5)	20 (5.8)	20 (2.5)	22 (2.5)	32 (2.88
Mood-stabilizing anticonvulsants	1169 (18.8)	86 (20.3)	48 (14.0)	149 (18.8)	129 (14.9)	219 (19.
Benzodiazepines	3413 (54.8)	272 (64.3)	179 (52.0)	412 (51.9)	474 (54.8)	621 (54.
Z-drugs	2846 (45.7)	263 (62.2)	184 (53.5)	365 (46.0)	422 (48.8)	542 (47.
History of self-poisoning, n (%)	1199 (19.2)	102 (24.1)	73 (21.2)	127 (16.0)	168 (19.4)	240 (21.
History of self-inflicting injuries, n (%)	248 (4.0)	25 (5.9)	6 (1.7)	24 (3.0)	34 (3.9)	49 (4.3)
Age-adjusted CCI, mean (SD)	0.7 (1.2)	0.8 (1.2)	0.8 (1.5)	0.8 (1.1)	0.8 (1.4)	0.7 (1.2)
Psychiatric service within 1 year before PD diagnosis, n (%)						
Psychiatric hospitalization	1910 (30.7)	137 (32.4)	112 (32.6)	225 (28.3)	268 (31.0)	379 (33.
Specialist psychiatric outpatient service	2564 (41.2)	161 (38.1)	132 (38.4)	342 (43.1)	335 (38.7)	448 (39.
Clinical psychology outpatient service	383 (6.1)	29 (6.9)	36 (10.5)	43 (5.4)	59 (6.8)	51 (4.5)
Community psychiatric service	670 (10.8)	46 (10.9)	34 (9.9)	88 (11.1)	90 (10.4)	132 (11.

Mood-stabilizing anticonvulsants included lamotrigine, carbamazepine, valproate. CCI=Charlson-comorbidity index; HKE=Hong Kong East Cluste
Central Cluster; KE=Kowloon East Cluster; KW=Kowloon West Cluster; NTE=New Territories East Cluster; NTW=New Territories West Cluster; I