

Childhood trauma and recent stressors in predicting subclinical psychotic symptoms among Chinese university students in southwest China: a machine learning analysis within a gender-specific framework

Wanjie Tang ¹, Zijian Deng,² Zeyuan Sun,³ Qijun Zhao,² Miguel Garcia-Argibay ^{4,5}, Kadan Anoop,⁶ Tao Hu,⁷ Shuang Xue,⁸ Natali Bozhilova,³ Aldo Conti,³ Steve Lukito,³ Siqi Wu ⁹, Gang Wang,¹⁰ Chunhan Jin,¹¹ Changjian Qiu,⁹ Qiaolan Liu,¹ Jay Pan,¹ Samuele Cortese ^{12,13}, Katya Rubia³

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjment-2025-301761>).

For numbered affiliations see end of article.

Correspondence to

Dr Wanjie Tang, West China School of Public Health, Sichuan University, Chengdu, China; tangwanjie2010@gmail.com
Professor Jay Pan; jay.pan@scu.edu.cn

WT and ZD are joint first authors.

SC and KR are joint senior authors.

Received 26 April 2025

Accepted 19 July 2025

ABSTRACT

Background Subclinical psychotic symptoms (SPS) are common among college students and can lead to future mental health issues. However, it is still not clear which specific childhood trauma, stressors and health factors lead to SPSs, partly due to confounding factors and multicollinearity.

Objective To use machine learning to find the main predictors of SPS among university students, with special attention to gender differences.

Methods A total of 21 208 university students were surveyed regarding SPS and a wide range of stress-related factors, including academic pressure, interpersonal difficulties and abuse. Nine machine learning models were used to predict SPS. We examined the relationship between SPS and individual stressors using χ^2 tests, multicollinearity analysis and Pearson heatmaps. Feature engineering, t-distributed stochastic neighborhood embedding (t-SNE) and Shapley Additive Explanation values helped identify the most important predictors. We also assessed calibration with calibration curves and Brier scores, and evaluated clinical usefulness with decision curve analysis (DCA) to provide a thorough assessment of the models. In addition, we validated this model using independent external data.

Findings The Extreme Gradient Boosting (XGBoost) model had the best prediction results, with an Area Under the Curve (AUC) of 0.89, and validated with external data. It also showed good calibration, and DCA indicated clear clinical benefit. Interpersonal difficulties, academic pressure and emotional abuse emerged as the strongest predictors of SPS. Gender-stratified analyses revealed that academic pressure and emotional abuse affected males more, while health issues like chest pain and menstrual pain were stronger predictors for females.

Conclusions Machine learning models effectively identified key stressors associated with SPS in university students. These findings highlight the importance of gender-sensitive approaches for the early detection and prevention of psychotic symptoms.

Clinical implications SPSs in college students can be predicted by interpersonal difficulties, academic stress and childhood emotional abuse. This information can

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The close associations between childhood trauma, stressful life events, health factors and subclinical psychotic symptoms (SPS) have already been established. However, it remains unclear which specific types of trauma and stressful events have the greatest impact. Comprehensive predictive models based on these factors have not yet been developed, especially for undergraduate populations.

WHAT THIS STUDY ADDS

⇒ By applying nine machine learning models, this study identified the optimal predictive model and found that interpersonal difficulties, academic stress and childhood emotional abuse are the three most influential factors for SPS. Additionally, gender differences were observed: males face higher risk mainly due to greater academic stress, while pain symptoms have a stronger impact on females. The model's performance was also validated using external data.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study provides a foundation for further optimisation and refinement of predictive models for assessing SPS risk in college students, as well as the development of targeted intervention strategies. It also offers clinical guidance for prevention and early intervention.

help mental health professionals develop better ways to prevent and address SPSs.

INTRODUCTION

Subclinical psychotic symptoms (SPS), such as thought control, paranoia and strange experiences, are prodromal symptoms of psychotic disorders.¹ It



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY. Published by BMJ Group.

To cite: Tang W, Deng Z, Sun Z, et al. *BMJ Ment Health* 2025;**28**:1–8.

is reported that the prevalence of SPS among the general population ranges from 7.2% to 26.69%.² Among individuals with SPS, 20% reported persistent symptoms, and about 7% eventually developed a psychotic disorder,³ which was associated with adverse outcomes. Additionally, SPS was associated with an increased risk of a variety of mental disorders beyond psychotic disorders, including affective, externalising and substance use disorders as well as poorer social function.⁴

Despite the existence of numerous aetiological models predicting SPS, they are sometimes constrained by limited sample sizes or an imbalance between clinical relevance and statistical significance. A growing body of evidence has suggested some potential risk factors such as childhood trauma, stressful life events, somatic symptoms and socioeconomic status, including pain symptoms and low family income.⁵ However, previous studies have neglected potential confounding effects and multicollinearity by considering risk factors in isolation, leading to possible misleading statistical results.⁶ To address this, a large-scale, data-driven study is needed to evaluate the combined impact of these factors, enabling more accurate predictions of SPS by considering their confounding effects rather than examining them independently.

Childhood trauma, including emotional, physical and sexual abuse, as well as neglect, is a well-established risk factor for SPS. Research has shown that exposure to trauma increases the likelihood of developing SPS.^{5,7} However, studies have shown mixed results about how different types of maltreatment affect SPS. For example, Rössler *et al*⁸ found links between SPS and physical or emotional abuse and neglect, but not sexual abuse.⁸ In contrast, Fekih-Romdhane *et al*⁹ found a link only with sexual abuse.⁹ Additionally, many existing studies have small sample sizes and other research limitations. This means we need larger, better-designed studies to better understand the link between childhood trauma and SPS.

According to the stress-vulnerability model,¹⁰ stress can trigger psychiatric symptoms when it exceeds an individual's coping capacity. For instance, a college study found significant associations between stressful life events and SPS,¹¹ while another study reported that adolescents with higher daily stress exhibited more prodromal psychotic symptoms.¹² However, some studies have found no link between life events and psychotic symptoms in young adults,¹³ highlighting the need for further investigation into the relationship between stressful life events and SPS.

Left-behind children (LBC) are also a potential risk factor affecting SPS. China once had tens of millions of LBC which refers to those in poor rural areas whose young parents leave home to work in cities, leaving their young children in the care of grandparents or other relatives.¹⁴ Furthermore, since Chinese culture often favours boys, stress may affect boys and girls in different ways. Therefore, this study will examine LBC and gender differences to better understand these issues. By identifying which childhood problems and stressful events have the greatest impact on SPS, we can offer better support to students in China.

Modern machine learning (ML) techniques have enabled researchers to examine multiple lines of risk factors simultaneously with increasing analytic rigour. ML techniques are superior to traditional statistical methods such as logistic regression in identifying risk factors in epidemiological studies in high dimensional data when there is poor a priori knowledge of the potential associations. Therefore, this study employed ML techniques on a comprehensive dataset of Chinese undergraduate students to identify potential risk factors for SPSs. The analysis took into account variables such as childhood trauma, recent

stressful life events and relevant sociodemographic and pain-related data.

METHODS

Participants and procedure

Participants were undergraduate students from three universities in Southwest China. Because China's enrolment policy assigns quotas to each province and ethnic group, the student body is nationally representative and diverse. Therefore, our sample includes students from many regions and ethnic backgrounds across the country.

Twenty research assistants and 150 assistant teachers were trained for 2 hours to help with the online survey. Participants could ask questions through WeChat or email and got quick answers. The survey link and informed consent form were shared in class WeChat groups. Participation was voluntary and students could withdraw at any time. After completing the survey, they received a small reward (1–3 RMB). To check data quality, we included simple attention questions, like 'What is the capital of China?' Surveys finished in less than 10 min were excluded.

A total of 31 602 students were invited to join the survey through WeChat groups. The online survey, run on the Wenjuanxing platform between March and April 2022, was completed by 21 534 students, giving a response rate of 68.1%. After removing 326 students who failed validity checks, the final sample included 21 208 students, with an average age of 19.71 years (range 16–30). To check for selection bias, we compared basic information (age, gender, grade and major) between participants and non-participants and found no significant differences (online supplemental table 1). This suggests selection bias was minimal.

Measures

Subclinical psychotic symptoms

SPS were measured over the past month using the *psychoticism* and *paranoid ideation* subscales of Symptom Checklist-90-R (SCL-90-R) Chinese version.¹⁵ Participants responded on a five-point Likert scale (0 = 'not at all', 1 = 'a little bit', 2 = 'moderately', 3 = 'quite a bit' and 4 = 'extremely often'). Categories of distress for both subscales were defined as follows: 'no distress' was indicated by a mean value of less than 1.00, while a mean value greater than 1.00 was classified as 'positive'. This classification follows the criteria used in previous studies.¹ Example items included: 'Someone else can control your thoughts', and 'Hearing voices others do not hear'. In the present study, Cronbach's α was 0.905 for the full SCL-16 SPS scale, 0.804 for the paranoid ideation subscale and 0.849 for the psychoticism subscale.

Stressful life events

Stressful life events were assessed using the modified Chinese Adolescent Self-Rating Life Events Checklist,¹⁶ a widely used tool with demonstrated reliability and validity for measuring stressful events in young adults. The 26-item scale covers five areas: interpersonal difficulties, academic pressure, punishment, personal loss and health/adaptability. In the current sample, Cronbach's α for the total scale was 0.892. Cronbach's α for the subscales also indicated moderate reliability: interpersonal difficulties, 0.778; academic pressure, 0.746; being punished, 0.679; personal loss, 0.693; and health and adaptability, 0.637.

Childhood trauma

Childhood trauma was assessed using the 28-item Childhood Trauma Questionnaire (CTQ),¹⁷ which measures five types of trauma experienced before age 16: emotional abuse, physical abuse, sexual abuse and emotional and physical neglect. Participants were asked to rate the childhood trauma on a 5-point scale (from 1=*never* to 5=*always*). In this study, the full scale had excellent reliability (Cronbach's $\alpha=0.864$), with satisfactory values for emotional abuse (0.765), physical abuse (0.762), sexual abuse (0.783), emotional neglect (0.942) and physical neglect (0.659).

Depressive symptoms

The Patient Health Questionnaire-9 (PHQ-9) is a widely used instrument designed to screen for the severity of depression. It has been validated in a Chinese population.¹⁸ The PHQ-9 is a nine-question tool used to check how often someone has experienced depressive symptoms in the past 2 weeks. Each question is scored from 0 to 3, with 0 meaning 'not at all' and 3 meaning 'nearly every day'. Total scores range from 0 to 27. Higher scores show worse depressive symptoms. In this study, the PHQ-9 demonstrated excellent reliability (Cronbach's $\alpha=0.854$).

Sociodemographic and health-related information

We gathered a range of factors based on previous research, including family history of mental illness, romantic relationships and LBC status. In addition, we collected demographic information such as age, gender, year of study and major. Health-related data were also collected, including information on chronic headaches and pain in the back, neck, chest and joints. Participants self-reported if they had frequently experienced any of these types of pain in the past year to an extent that interfered with their work or studies.

Statistical analyses

All statistical analyses were performed using SPSS V.22.0 (IBM, Chicago, Illinois, USA) for descriptive statistics, and Python V.3.9.17 with sklearn V.1.3.0 for ML methods. Descriptive statistics were used to summarise sociodemographic and clinical information. χ^2 tests checked for differences in SPSs across sociodemographic groups, childhood abuse types and stressful life events. We calculated the Variance Inflation Factor (VIF) to check for multicollinearity among the independent variables. A Pearson correlation heatmap was also created to visually show how the independent variables are related to each other.

The nine ML models were chosen based on experience from previous model selection in the literature¹⁹: Gradient Boosting Tree, Logistic Regression, Random Forest, Support Vector Machine, Multilayer Perceptron, the Extreme Gradient Boosting (XGBoost), Naive Bayes, k-nearest neighbours, CatBoost, AdaBoost (see online supplemental section 1 for details). By evaluating these models, we aim to identify the optimal methodological approach. Participants were randomly divided into training (80%) and testing (20%) sets for each fold. Model performance was evaluated across folds using metrics such as average AUC, balanced accuracy and Area Under the Precision-Recall Curve.

The performance of the optimal ML model was evaluated using Receiver Operating Characteristic (ROC) and P-R curves, along with metrics such as sensitivity (Sen), specificity (Spe), positive predictive value (PPV) and negative predictive value (NPV). We also performed a decision curve analysis for the XGBoost model to assess clinical utility, and calculated the calibration curve and Brier score to evaluate the model's calibration performance. To

further explore the impact of individual variables on SPS, we applied Shapley Additive Explanation (SHAP) to quantify the contribution of each feature to the model's predictions. SHAP waterfall plots were used to provide detailed insights into how individual features influenced predictions for specific samples. Additionally, feature engineering and t-SNE visualisation were employed to effectively display how features distinguished positive and negative samples in 2D space.

A key aspect of our analysis was evaluating algorithmic fairness and potential gender bias. To further investigate the relationship between sexual abuse and SPS, we used a mediation model to examine how depression mediates this relationship.

To validate the performance of our prediction model, we will employ an independent, non-public dataset comprising approximately 8000 university students recruited from universities in southern and southwestern China.

RESULTS

Clinical features, multicollinearity assessment and χ^2 test results

Approximately 15% of participants reported experiencing SPS, with 7.4% endorsing paranoid ideation (mean (SD): 2.42 (3.68)) and 13.8% endorsing psychoticism (mean (SD): 1.65 (2.47)).

The heatmap confirmed generally low inter-variable correlations (online supplemental figure 1), with only a few pairs showing moderate to high correlation coefficients ($|r|>0.8$). Furthermore, the multicollinearity assessment showed that most variables had VIF values below 2 (online supplemental figure 2), with the highest being physical abuse (VIF=4.25), which is still well below the commonly accepted thresholds (VIF >5 or 10). Overall, these results show that multicollinearity is not a major concern in our study.

χ^2 tests showed that students with a history of psychological or sexual abuse were more likely to have SPS (see online supplemental table 2). SPS were also more common in students with high academic pressure and interpersonal difficulties. We used False Discovery Rate correction to adjust p values for multiple comparisons.

ML results

XGBoost outperformed the other ML model (online supplemental table 3), achieving the highest AUC (0.89, 95% CI: 0.88 to 0.90), balanced accuracy (0.80) (figure 1) (Section 1 of the online supplemental appendix lists the nine ML methods, their settings and evaluation metrics. Section 2 provides the rationale for not adopting Lasso for feature selection in this study, as shown in online supplemental figure 3), and shows the missing data pattern in online supplemental figure 4). It also had strong discriminatory power (online supplemental figure 5) for t-SNE visualisation.

The XGBoost model performed well at two thresholds (see figure 2). At a threshold of 0.17 (chosen for best balance of sensitivity and specificity), the model had 85.76% sensitivity, 79.10% specificity, 96.98% NPV and 41.45% PPV. At a threshold of 0.76, where NPV and PPV were balanced, the model showed a specificity of 98.09%, sensitivity of 33.60%, NPV of 89.53% and PPV of 75.00%. More threshold results are in online supplemental table 4.

The calibration curve was close to the ideal line, showing the model's predictions were mostly accurate, with only a slight overestimation (see figure 2c). The model also gave a higher net benefit than always treating everyone at most thresholds, meaning it helped improve decisions and reduce unnecessary

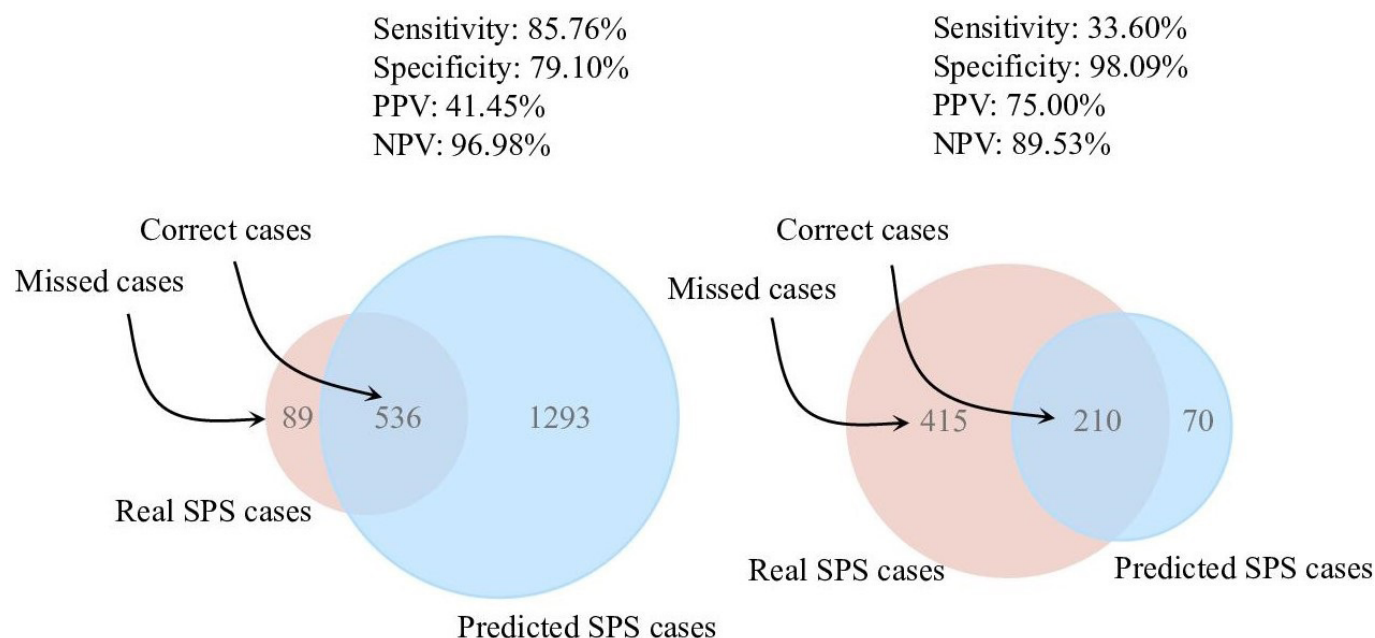


Figure 1 Results of XGBoost on test set. NPV, negative predictive value; PPV, positive predictive value.

treatments. The red dot marks the threshold with the highest net benefit compared with treating all patients, which is useful for doctors (see [figure 2d](#)). Overall, these results show the model performed very well.

We tested the model on a separate external SPS dataset to see how well it works on new data. The XGBoost model got an AUC of 0.8865 and a balanced accuracy of 0.7752 (online supplemental figure 6), showing it is accurate and generalises well.

Online supplemental figure 7 shows the ROC curves for different k values of Synthetic Minority Over-sampling Technique, and the model performed well in all cases, showing stable and reliable results.

Feature importance analysis

[Figure 3a](#) shows the 20 most important features for the model, ranked by SHAP values. Interpersonal difficulties were the strongest predictor of SPS, followed by academic pressure. Emotional abuse also contributed but was less important than the top two factors. Section 3 of the online supplemental appendix explains how we used SHAP to find important features and checked these results with other methods to make sure they are reliable. Online supplemental table 5 provides feature importance rankings from traditional methods. SHAP waterfall plots (online supplemental figure 8) show how each factor raises or lowers the predicted risk of SPS.

Both Gain and SHAP found that interpersonal difficulties and academic pressure are the main predictors of SPSs. However, SHAP also showed how less important features, like pain-related variables, can still strongly affect some individuals' risk. This means that while some factors are important overall, other features may matter more for certain people.

Gender differences in ML models

Gender analysis showed that different factors affect SPS in males and females ([figure 3b](#)). Academic pressure and emotional abuse had a bigger impact on males. In females, health issues like chest pain and dysmenorrhoea were more important. Interpersonal difficulties and academic pressure were top predictors for both

genders, but their influence varied. These results suggest that gender should be considered when interpreting the model.

To better measure how gender influences SPS risk and reduce potential confounding, we used a Causal Forest model. The right-skewed distribution of Individual Treatment Effects suggests that gender impacts SPS risk for some individuals. The Average Treatment Effect was 0.04, indicating a slightly higher SPS risk in males after adjusting for other factors. This analysis offers a more detailed understanding of how gender moderates the effects of different risk factors (online supplemental figure 9).

Mediation model

The analysis reveals a significant indirect effect ($\beta=0.11$, $p<0.001$), indicating that depressive symptoms mediate the relationship between sexual abuse and SPS. This finding supports the hypothesis that sexual abuse affects SPS both directly and indirectly through its impact on depressive symptoms (see online supplemental table 6 and figure 10).

DISCUSSION

Our ML models revealed strong associations between stressful life events, childhood trauma and SPS in undergraduates. This study, the first of its kind to use ML on a large sample to assess multiple SPS risk factors, confirmed established links such as abuse, social adversity and physical health issues. Notably, we identified two primary risk factors: interpersonal difficulties and academic stress, which significantly impacted SPS. Emotional abuse, though less influential, still contributed to SPS, highlighting the importance of addressing this often-overlooked form of maltreatment. Specifically, gender differences were also observed in the model. Academic pressure and emotional abuse had a stronger effect on males, while physical health issues like chest pain and dysmenorrhoea were more significant for females. These findings underscore the distinct ways in which various factors affect males and females.

This study underscores the critical role of interpersonal difficulties in the development of SPS, building on prior research

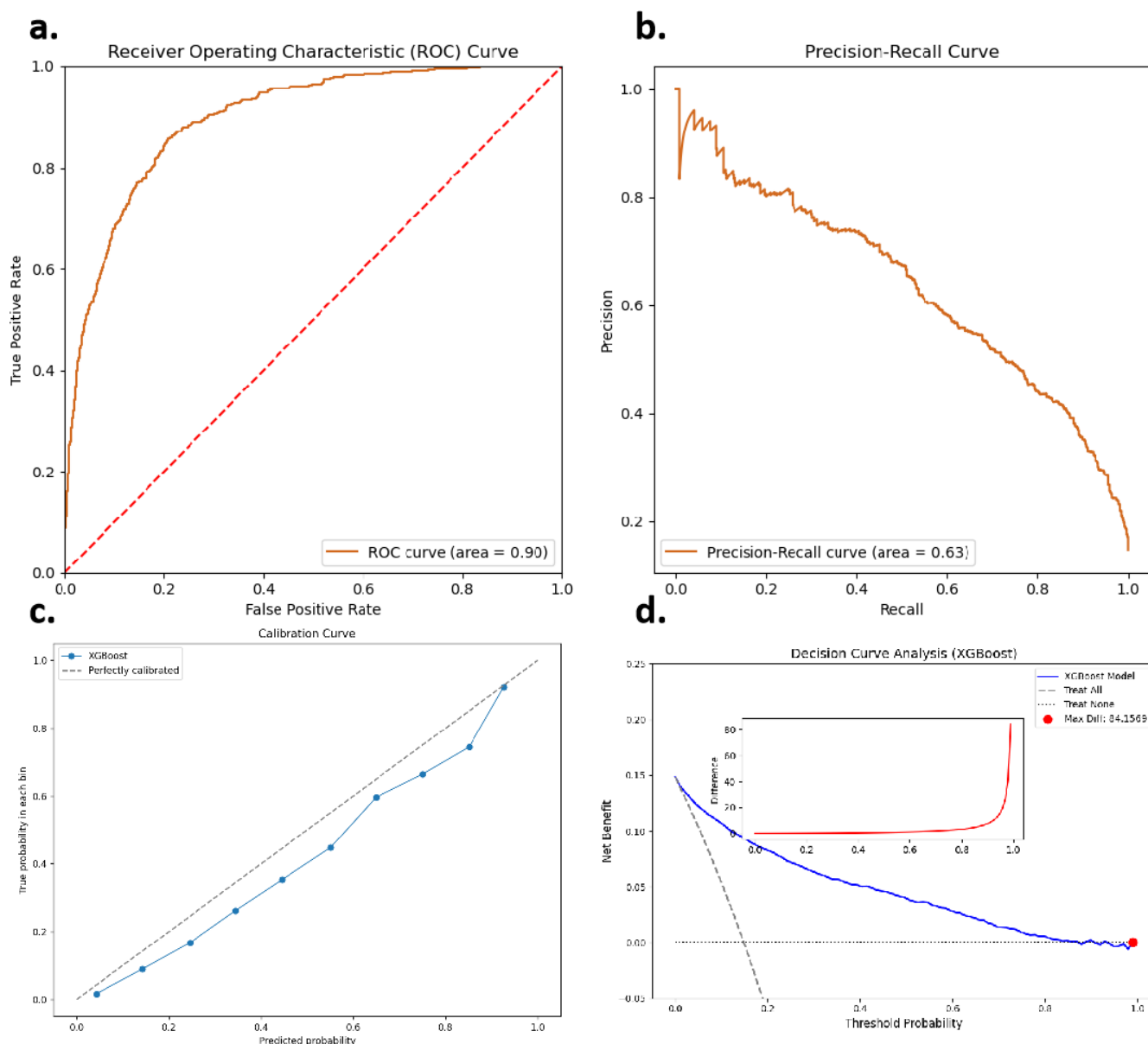


Figure 2 Depiction of the XGBoost performance predicting subclinical psychotic symptoms at two different thresholds. The probability threshold was 0.17 (left) and 0.76 (right). (A) Receiver Operating Characteristic curve, (B) precision-recall curve, (C) calibration curve and (D) decision curve.

that found young people with SPS often report more interpersonal challenges during adolescence, emphasising the significant impact of personal difficulties on SPS.²⁰ Longitudinal twin cohort studies also highlight the protective role of interpersonal relationships in reducing the impact of poly-victimisation on SPS.²¹ Our findings align with a semistructured interview study, which showed that individuals at risk for psychosis often face significant interpersonal challenges.²² However, our study extends these findings by demonstrating that interpersonal difficulties are the most significant factors among various adverse life stressors and childhood traumas contributing to SPS. Studies suggest that psychosocial interventions focusing on interpersonal skills can aid rehabilitation in individuals with psychotic disorders,²³ supporting our conclusion that addressing interpersonal difficulties is crucial in preventing SPS. Therefore, targeted psychological and social interventions addressing interpersonal challenges may help protect young people from SPS and its long-term consequences.

In line with previous research,²⁴ we also observed that academic stress was another significant contributor to SPS among undergraduates. In the current education system in Mainland China, exam-performance-driven progression reaches from primary school to postgraduate studies. Young people have been facing increasingly fierce competition and peer pressure since childhood, and most likely throughout their careers. Interestingly, a cross-cultural study showed that the academic burden was highest in students from China who were the most stressed compared with students from the USA, Japan and South Korea.²⁵ Such academic pressure on students in China may originate from their high-standard requirements, parents' expectations and social values of praising the 'tiger parenting'. These internal and external pressures stemming from intense competition can increase college students' stress levels, lead to burnout and heighten their vulnerability to mental health problems. Our findings highlight academic stress as an important, yet previously overlooked, target for practitioners aiming to prevent or intervene in SPS among college students.

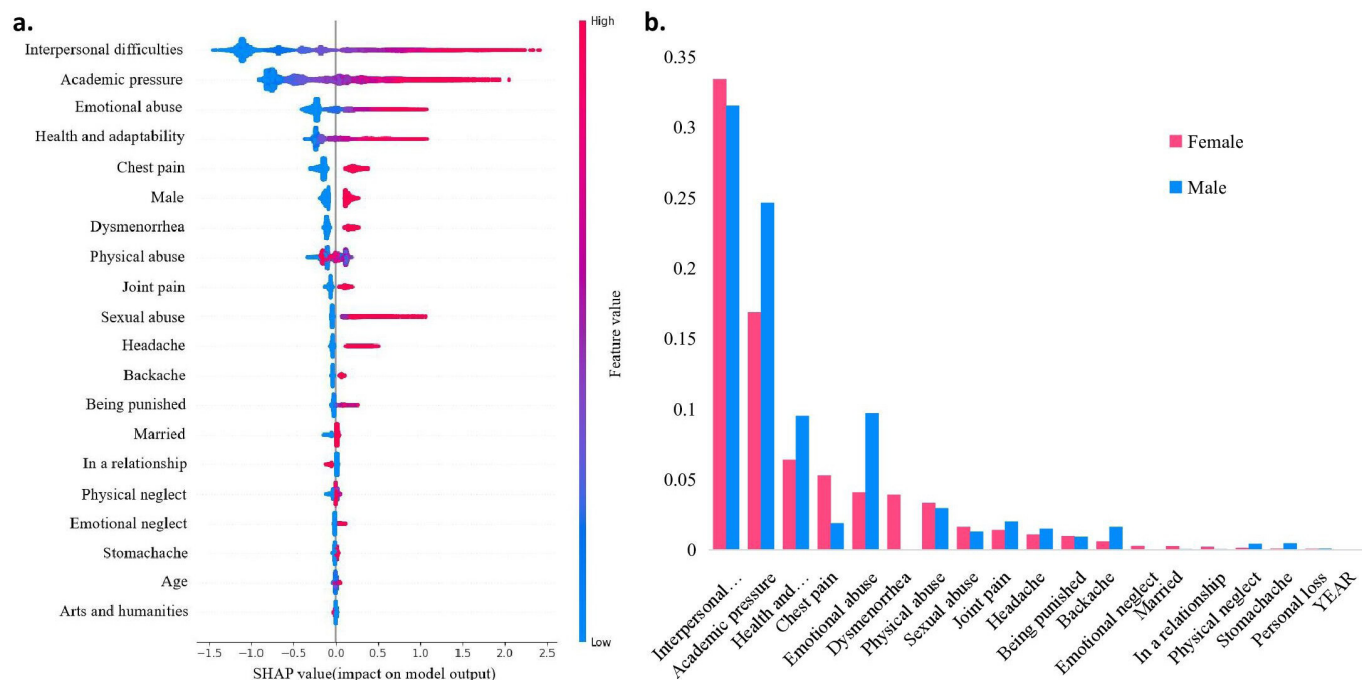


Figure 3 Summary SHAP plot. (A) Feature importance for SPS classification calculated by SHAP. (B) Influence of various features across gender. SHAP, Shapley Additive Explanation; SPS, subclinical psychotic symptom.

In our data, we found that emotional abuse, more than other forms of childhood trauma, is independently associated with SPS in this population. Emotional abuse, often under-recognised, refers to harmful interactions in the carer–child relationship that do not involve physical contact. Our results align with prior studies that emotional abuse was the only form of childhood trauma associated with SPS, possibly because emotional abuse occurs more frequently than other types of abuse subtypes,²⁶ increasing the risk of SPS. From a neurodevelopmental perspective, emotional abuse can disrupt an individual's hypothalamic–pituitary–adrenal axis,²⁷ increasing the risk of developing SPS.

While our data show a significant association between sexual abuse and SPS, this effect diminished when other risk factors, such as academic stress and interpersonal difficulties, were included in the ML models. Although academic stress, interpersonal difficulties and emotional abuse are the primary contributors, the role of sexual abuse in SPS should not be overlooked.²⁸ This may be because our analysis showed that depressive symptoms connect sexual abuse and SPS.

Our findings reveal gender-specific differences in the impact of various factors on SPS among undergraduates, offering new insights into the complexity of risk factors for males and females. Both academic pressure and emotional abuse emerged as prominent predictors for both genders, but these factors had a stronger influence on males, particularly academic pressure. This suggests that boys may experience greater stress and emotional abuse due to the high expectations placed on them in Chinese culture. Traditionally, men are expected to shoulder more responsibility, and society tends to look down on men who are unsuccessful in their careers as well as women who display excessive career ambition. For example, a study shows that in Chinese academia, men are often expected to achieve more and support their families, while less is expected of women.²⁹ Therefore, cultural beliefs and traditional gender roles put extra pressure on boys, making these problems worse. These factors show that culture affects stress differently for male and female. For females, health

issues like chest pain and period pain increase their risk of SPS. This means it's important to pay attention to gender-specific health problems in future studies.

These gender differences show that we need different support strategies for male and female students. Future studies should pay attention to how gender affects SPS and why some issues, like academic pressure, may affect males more. Understanding the social, psychological and biological factors behind these differences in the future will help us develop more effective support programmes for all undergraduate students.

Despite using large samples and robust ML methods to explore risk factors for SPS among undergraduate students in Southwest China, several limitations must be considered when interpreting the results. First, assessments of childhood trauma and stress events rely on retrospective self-reports, which may be influenced by recall bias or social desirability effects. Future studies should incorporate both objective and subjective measures to strengthen the findings and provide further validation. Second, we did not assess substance use or mood/anxiety symptoms, which could be important covariates or mediators; these were excluded based on pilot study findings and expert input. Third, due to sensitivity concerns, mental illness diagnoses were not collected. Fourth, while we excluded participants who failed comprehension checks, some may have misinterpreted questions. Fifth, accuracy differences across the nine ML models were minimal, but the use of only XGBoost results may reflect model bias. Sixth, pain symptoms had low predictive value and were not retained in the final prediction model; this may limit the clinical applicability of our model in populations where such symptoms are common, such as individuals with chronic health conditions. Finally, as our study focused on university students from Southwest China within a limited age range, and given the relatively high non-response rate, there may be some bias in our results. Therefore, our findings should be interpreted with caution when considering their generalisability to university students nationwide or all Chinese youth. Furthermore, our model still requires further

research and external validation to ensure its reliability and wider application.

Previous studies in Western countries have found that most people with SPS experienced childhood trauma or negative life events,³⁰ but they did not build predictive models for SPS based on these factors, so our study offers new insights into what could affect SPS in this population. This can help future research and support programmes for students.

In conclusion, our study identified key risk factors for SPS in university students, using ML and large sample sizes to assess both childhood trauma and recent stressors. We confirmed known risk factors like domestic abuse and social adversity, but highlighted interpersonal difficulties and academic stress and emotional abuse as the most significant contributors. Gender differences were observed, with academic pressure and emotional abuse having a stronger impact on males, while health issues like chest pain and dysmenorrhoea were more significant for females, offering a gender-differentiated perspective on SPS risk and suggesting the need for gender-tailored interventions. Future studies should explore the mechanisms behind gender differences, track SPS risk factors longitudinally and include additional variables like substance use and mood/anxiety symptoms. Expanding this research across different cultural contexts will further refine these findings and improve intervention strategies.

It is also important to consider how cultural and psychosocial stressors may interact with neurodevelopmental pathways to increase the risk for SPS. Ongoing interpersonal difficulties, academic stress and emotional abuse within a high-pressure environment may disrupt the HPA axis,²⁷ which links early stress to later mental health symptoms. This helps provide a biological explanation for our findings and suggests that both cultural and biological factors should be considered in future research. However, the specific ways cultural stress affects neurobiology and gender differences still need more study.

Author affiliations

¹West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China

²School of Computing Science, Sichuan University, Chengdu, China

³Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

⁴School of Medical Sciences, Örebro universitet, Örebro, Sweden

⁵Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

⁶University of Southampton, Southampton, UK

⁷Chengdu Normal University, Chengdu, Sichuan, China

⁸Norwegian University of Science and Technology, Trondheim, Trøndelag, Norway

⁹West China Hospital of Medicine, Chengdu, Sichuan, China

¹⁰Chengdu University, Chengdu, Sichuan, China

¹¹Chongqing Jiaotong University, Chongqing, Chongqing, China

¹²University of Southampton, Southampton, UK

¹³Academic Unit of Psychology and Clinical and Experimental Sciences (CNS and Psychiatry), University of Southampton, Southampton, UK

Acknowledgements We thank all the collaborators in our research, especially Professor Helen Fisher from King's College London for her help and support.

Contributors WT and ZD share first authorship of this study. WT, ZD and JP verified and accessed the data used in the study. WT and JP served as guarantors. WT and JP are joint corresponding authors. WT was responsible for conceptualisation, data curation, formal analysis, investigation, methodology, visualisation, writing—original draft and writing—review and editing. ZD conducted data analyses for the Methods section, prepared figures and tables and contributed to drafting and reviewing the manuscript. ZS contributed to figure preparation, methodology and writing—review and editing. QZ, MG-A and KA contributed to the implementation and interpretation of machine learning methods, provided methodological guidance and participated in writing—review and editing. TH, SX, SW, GW, CJ, CQ and QL were involved in data investigation and contributed to writing—review and editing. NB, AC and SL contributed to conceptualisation, methodology and writing—review and editing.

JP, SC and KR provided funding support. JP was responsible for conceptualisation, formal analysis, investigation, methodology, visualisation, supervision and writing—review and editing. KR and SC are joint senior authors, and provided ongoing supervision and guidance throughout the study. All the authors reviewed and approved submission of the final version of the manuscript.

Funding SC is funded by the NIHR Research Professor (NIHR303122) for this research. The views expressed are those of the authors and not necessarily those of the NIHR, NHS or the UK Department of Health and Social Care. SC is also supported by several other NIHR grants (NIHR203684, NIHR203035, NIHR130077, NIHR128472, RP-PG-0618-20003) and by grant 101095568-HORIZONHLTH-2022-DISEASE-07-03 from the European Research Executive Agency. AC, NB, SC and KR were supported by the Efficacy and Mechanism Evaluation (EME) programme, a partnership between the MRC and NIHR (project ref: NIHR130077), SC and KR by the NIHR Programme Grant (NIHR203684) and SC by additional NIHR grants. KR is supported by the National Institute for Health Research (NIHR) and the UK Department of Health via the NIHR Biomedical Research Centre for Mental Health at South London and the Maudsley NHS Foundation Trust and King's College London (no grant number; not applicable). The design, management, analysis and reporting of the study are independent of the funders. WT was supported by the project of the Institute of New Productive Forces for Health, West China School of Public Health/ West China Fourth Hospital, Sichuan University (Project No. HN240101A).

Competing interests KR has received funding from TAKEDA for a separate project. SC has declared reimbursement for travel and accommodation expenses from the Association for Child and Adolescent Central Health (ACAMH), the Canadian ADHD Alliance Resource, the British Association of Psychopharmacology, Healthcare Convention and CCM Group for educational activities on ADHD, and has received honoraria from Medice. The authors declare that they have no further conflicts of interest.

Patient consent for publication Consent obtained directly from patient(s).

Ethics approval The study was approved by the Ethics Committee of the Sichuan Psychological Association (NO. SCXL202209). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The data used in this study, as well as the independent external dataset used for validation, are available from the corresponding author upon request. The variable dictionary, main analysis code and data preprocessing procedures are available on GitHub (<https://github.com/Diciembre08/SPS-analysis>).

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iDs

Wanjie Tang <http://orcid.org/0000-0002-1307-6608>

Miguel Garcia-Argibay <http://orcid.org/0000-0002-4811-2330>

Siqi Wu <http://orcid.org/0009-0009-1546-4696>

Samuele Cortese <http://orcid.org/0000-0001-5877-8075>

REFERENCES

- 1 Rössler W, Ajdacic-Gross V, Haker H, *et al*. Subclinical psychosis syndromes in the general population: results from a large-scale epidemiological survey among residents of the canton of Zurich, Switzerland. *Epidemiol Psychiatr Sci* 2015;24:69–77.
- 2 Zammit S, Kounali D, Cannon M, *et al*. Psychotic experiences and psychotic disorders at age 18 in relation to psychotic experiences at age 12 in a longitudinal population-based cohort study. *Am J Psychiatry* 2013;170:742–50.
- 3 Linscott RJ, van Os J. An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: on the pathway from proneness to persistence to dimensional expression across mental disorders. *Psychol Med* 2013;43:1133–49.
- 4 Solmi M, Soardo L, Kaur S, *et al*. Meta-analytic prevalence of comorbid mental disorders in individuals at clinical high risk of psychosis: the case for transdiagnostic assessment. *Mol Psychiatry* 2023;28:2291–300.

- 5 Fisher HL, Jones PB, Fearon P, *et al.* The varying impact of type, timing and frequency of exposure to childhood adversity on its association with adult psychotic disorder. *Psychol Med* 2010;40:1967–78.
- 6 Vatcheva KP, Lee M, McCormick JB, *et al.* Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. *Epidemiology (Sunnyvale)* 2016;6:227.
- 7 Croft J, Heron J, Teufel C, *et al.* Association of Trauma Type, Age of Exposure, and Frequency in Childhood and Adolescence With Psychotic Experiences in Early Adulthood. *JAMA Psychiatry* 2019;76:79–86.
- 8 Rössler W, Ajdacic-Gross V, Rodgers S, *et al.* Childhood trauma as a risk factor for the onset of subclinical psychotic experiences: Exploring the mediating effect of stress sensitivity in a cross-sectional epidemiological community study. *Schizophr Res* 2016;172:46–53.
- 9 Fekih-Romdhane F, Tira S, Cheour M. Childhood sexual abuse as a potential predictor of psychotic like experiences in Tunisian college students. *Psychiatry Res* 2019;275:181–8.
- 10 Zubin J, Spring B. Vulnerability—a new view of schizophrenia. *J Abnorm Psychol* 1977;86:103:103–26.
- 11 Ered A, Gibson LE, Maxwell SD, *et al.* Coping as a mediator of stress and psychotic-like experiences. *Eur Psychiatry* 2017;43:9–13.
- 12 Tessner KD, Mittal V, Walker EF. Longitudinal study of stressful life events and daily stressors among adolescents at high risk for psychotic disorders. *Schizophr Bull* 2011;37:432–41.
- 13 Thompson KN, Phillips LJ, Komesaroff P, *et al.* Stress and HPA-axis functioning in young people at ultra high risk for psychosis. *J Psychiatr Res* 2007;41:561–9.
- 14 Fellmeth G, Rose-Clarke K, Zhao C, *et al.* Health impacts of parental migration on left-behind children and adolescents: a systematic review and meta-analysis. *The Lancet* 2018;392:2567–82.
- 15 Yu Y, Wan C, Zhao X, *et al.* Undergraduate students' norms for the Chinese version of the symptom check-List-90-R (SCL-90-R). *BMC Public Health* 2020;20:1588.
- 16 Liu X, Liu L, Yang J, *et al.* The Adolescent Self-Rating Life Events Checklist and its reliability and validity. *Chinese Journal of Clinical Psychology* 1997;5:34–6.
- 17 Bernstein DP, Stein JA, Newcomb MD, *et al.* Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse & Neglect* 2003;27:169–90.
- 18 Wang W, Bian Q, Zhao Y, *et al.* Reliability and validity of the Chinese version of the Patient Health Questionnaire (PHQ-9) in the general population. *Gen Hosp Psychiatry* 2014;36:539–44.
- 19 Raschka S. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv Preprint arXiv* 2018.
- 20 Coughlan H, Walton-Ball E, Carey E, *et al.* Self-reported interpersonal and educational/vocational difficulties in young adults with a history of transient psychotic experiences: findings from a population-based study. *BMC Psychiatry* 2021;21:30.
- 21 Crush E, Arseneault L, Danese A, *et al.* Using discordant twin methods to investigate an environmentally mediated pathway between social support and the reduced likelihood of adolescent psychotic experiences. *Psychol Med* 2020;50:1898–905.
- 22 MacBeth A, Schwannauer M, Gumley A. The association between attachment style, social mentalities, and paranoid ideation: An analogue study. *Psychol Psychother* 2008;81:79–93.
- 23 Lloyd C, Sullivan D, Williams PL. Perceptions of social stigma and its effect on interpersonal relationships of young males who experience a psychotic disorder. *Aus Occup Therapy J* 2005;52:243–50.
- 24 Coid JW, Zhang Y, Sun H, *et al.* Impact of urban birth and upbringing on expression of psychosis in a Chinese undergraduate population. *BMC Psychiatry* 2021;21:493.
- 25 Evening B. Survey among Chinese, Japanese, Korean, and American high school students shows that Chinese students experience the highest academic pressure. *Research in Educational Development* 2010;7:21.
- 26 Tschoeke S, Flammer E, Bichescu-Burian D, *et al.* The Association between Type of Dissociation and Psychotic Experiences in a Non-Psychotic Inpatient Sample. *Journal of Trauma & Dissociation* 2022;23:504–20.
- 27 Marques-Feixa L, Palma-Gudiel H, Romero S, *et al.* Childhood maltreatment disrupts HPA-axis activity under basal and stress conditions in a dose-response relationship in children and adolescents. *Psychol Med* 2023;53:1060–73.
- 28 Kilcommons AM, Morrison AP, Knight A, *et al.* Psychotic experiences in people who have been sexually assaulted. *Soc Psychiatry Psychiatr Epidemiol* 2008;43:602–11.
- 29 Horta H, Tang L. Male and female academics' gendered perceptions of academic work and career progression in China. *Higher Education Quarterly* 2023;77:515–36.
- 30 Paetzold I, Myin-Germeys I, Schick A, *et al.* Stress reactivity as a putative mechanism linking childhood trauma with clinical outcomes in individuals at ultra-high-risk for psychosis: Findings from the EU-GEI High Risk Study. *Epidemiol Psychiatr Sci* 2021;30:e40.