


# BMJ Open Ningbo Schizophrenia Cohort (NSC) – a longitudinal ambispective cohort based on electronic health records: cohort profile

He Ye,<sup>1</sup> Yang Zhao,<sup>2</sup> Lian Li,<sup>3,4</sup> Yi Qian,<sup>1</sup> Hangjie Zhu,<sup>1</sup> Guolin Bian,<sup>3,4</sup> Liya Liu <sup>1,3,4</sup>

**To cite:** Ye H, Zhao Y, Li L, et al. Ningbo Schizophrenia Cohort (NSC)—a longitudinal ambispective cohort based on electronic health records: cohort profile. *BMJ Open* 2025;**15**:e091188. doi:10.1136/bmjopen-2024-091188

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2024-091188>).

Received 15 July 2024  
Accepted 30 April 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

<sup>1</sup>School of Public Health, Ningbo University Health Science Center, Ningbo, Zhejiang, China

<sup>2</sup>Department of Biostatistics, Nanjing Medical University, Nanjing, Jiangsu, China

<sup>3</sup>Department of Psychiatry, Ningbo Kangning Hospital, Ningbo, Zhejiang, China

<sup>4</sup>Department of Psychiatry, Affiliated Kangning Hospital of Ningbo University, Ningbo, Zhejiang, China

## Correspondence to

Dr Liya Liu;  
[liuliya@nbu.edu.cn](mailto:liuliya@nbu.edu.cn) and  
Dr Guolin Bian;  
[bianguolin2025@163.com](mailto:bianguolin2025@163.com)

## ABSTRACT

**Purpose** Electronic health records (EHR) passively generate large datasets on real-world patient populations in easily retrievable form, allowing the cost-efficient and timely execution of epidemiological cohorts on a broad array of topics. However, EHR-based cohorts specialising in mental disorders have not yet been reported. Ningbo has made significant achievements in healthcare data management in China. This study, relying on the Ningbo Mental Health Information System and the Ningbo Regional Health Information Platform, has established the Ningbo Schizophrenia Cohort (NSC), providing an exemplary study for cohort studies on schizophrenia.

**Participants** This population-based ambispective cohort study included patients with schizophrenia aged 18–65 years at the time of diagnosis who were eligible for healthcare services in Ningbo, China. Participants were identified using the Ningbo Mental Health Information System between 1 January 2010, and 31 December 2023. Once an individual enters the NSC, they are followed up continuously until death or relocation. A total of 26 899 patients with schizophrenia are included in the NSC.

**Findings to date** Among 26 899 patients, 55.4% were female and 53.1% had less than 7 years of education. Until 31 December 2023, 4505 deaths occurred, and 97.83% of patients had at least one electronic medical record. The median age at diagnosis for non-survivors (median (IQR): 40 (29–51) years) was higher than that of survivors (median (IQR): 34 (26–45) years).

**Future plans** The NSC will continue to collect longitudinal data to capture the full life cycle of schizophrenia, including pre-onset, diagnosis, follow-up, recovery or death. This will result in a continuous, complete and multidimensional EHR for patients with schizophrenia. Planned future research aims to generate new real-world evidence on the aetiology of schizophrenia, investigate comorbidities to facilitate co-management and develop predictive models for schizophrenia and related cardiovascular diseases.

**Trial registration number** [NCT06370793](https://doi.org/10.1136/bmjopen-2024-091188).

## INTRODUCTION

The 2021 Global Burden of Diseases, Injuries and Risk Factors Study revealed that 13.9% of the global population has experienced

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Ningbo Schizophrenia Cohort (NSC) provides ambispective, observational and multicentre cohort data collected from a population of patients with schizophrenia.
- ⇒ NSC is fully linked to electronic health records, and its study design involves collecting data across the full life cycle of patients.
- ⇒ NSC has a large sample size and broad coverage, collecting data from most patients with schizophrenia in Ningbo.
- ⇒ NSC may have potential information bias in this study.

mental disorders, which are among the top 10 causes of this global burden.<sup>1</sup> The estimated number of cases of mental disorders rose from 654.8 million in 1990 to 1.1 billion in 2021, exhibiting a 68% increase.<sup>2,3</sup> Among non-communicable diseases (NCDs), mental disorders account for 28% of disability-adjusted life years and have a greater impact on health than cardiovascular disease (CVD) or cancer.<sup>4,5</sup> Schizophrenia is a chronic disorder characterised by positive symptoms (hallucinations and delusions), negative symptoms (blunted emotional responses and impaired motivation) and cognitive impairments.<sup>6</sup> The cause of the disease remains unclear, with about 23.2 million people worldwide suffering from schizophrenia, which indicates a prevalence rate of about 0.28%.<sup>7</sup> In China, the weighted lifetime prevalence of schizophrenia is 0.7%.<sup>8</sup> Patients with schizophrenia have a poor prognosis and their mortality rate is 2–3 times higher than that of the general population.<sup>9</sup> Early diagnosis and intervention in patients is important to improve prognosis,<sup>10</sup> which requires urgent improvement.

While classic prospective cohort studies, such as the Framingham Heart Study,<sup>11</sup> have

pioneered modern aetiological research, building and maintaining a large cohort is costly and challenging.<sup>12</sup> Electronic health records (EHR) provide an important source of real-world data that can address various research questions, including disease aetiology, characterisation and prognosis evaluation.<sup>13</sup> The extensive data within EHR systems offer a cost-effective and rich resource of structured and unstructured healthcare information.<sup>14 15</sup> EHR-based cohorts have been established to cover a wide range of research areas on NCDs, including chronic kidney disease,<sup>16</sup> chronic cough<sup>17</sup> and CVD.<sup>18 19</sup> However, schizophrenia as an important chronic disease has not been reported in EHR-based specialty cohort studies. The past 3 years have witnessed a gradual increase in the reporting of studies such as risk factor identification and diagnostic/prognostic modelling for mental disorders based on EHR.<sup>20 21</sup> This indicates the feasibility and validity of generating high-quality evidence in the field of mental disorders research based on EHR. Therefore, we use the strengths of the Ningbo Mental Health Information System (NMHIS) and Ningbo Regional Health Information Platform (NRHIP) to conduct a cohort study of patients with schizophrenia in China.

The Ningbo Schizophrenia Cohort (NSC) is a specialised EHR-based cohort of patients with schizophrenia that can provide real-world evidence to infer aetiology and enhance patient prognosis. It is also useful for improving mental health services and enhancing patients' quality of life by obtaining information such as health service utilisation. The utilisation of NSC includes but is not limited to (1) understanding the epidemiological characteristics of schizophrenia, (2) inferring the aetiology of schizophrenia, (3) improving the prognosis of patients with schizophrenia and decreasing the risk of disease recurrence or readmission, (4) conducting comorbidity studies of schizophrenia and (5) providing guidance on risk stratification for serious comorbidity in patients with schizophrenia.

## COHORT DESCRIPTION

### Study area

Ningbo is located in the southern part of China's Yangtze River Delta, covering an area of approximately 9816 km<sup>2</sup>. It features a mild and humid subtropical monsoon climate with four distinct seasons. In 2024, the gross domestic product of Ningbo, with a total resident population of 9 777 000 individuals, reached ¥1.81 trillion.<sup>22 23</sup> The city has 225 medical institutions, including 8 tertiary hospitals, 52 secondary hospitals and 165 primary healthcare institutions. Secondary and tertiary hospitals primarily deliver diagnostic services, outpatient care and inpatient treatment, while primary healthcare institutions, such as community health centres and township clinics, focus on continuity of follow-up care.

Ningbo was selected for this study for two primary reasons: (1) Ningbo offers a comprehensive range of mental health primary healthcare institutions with a

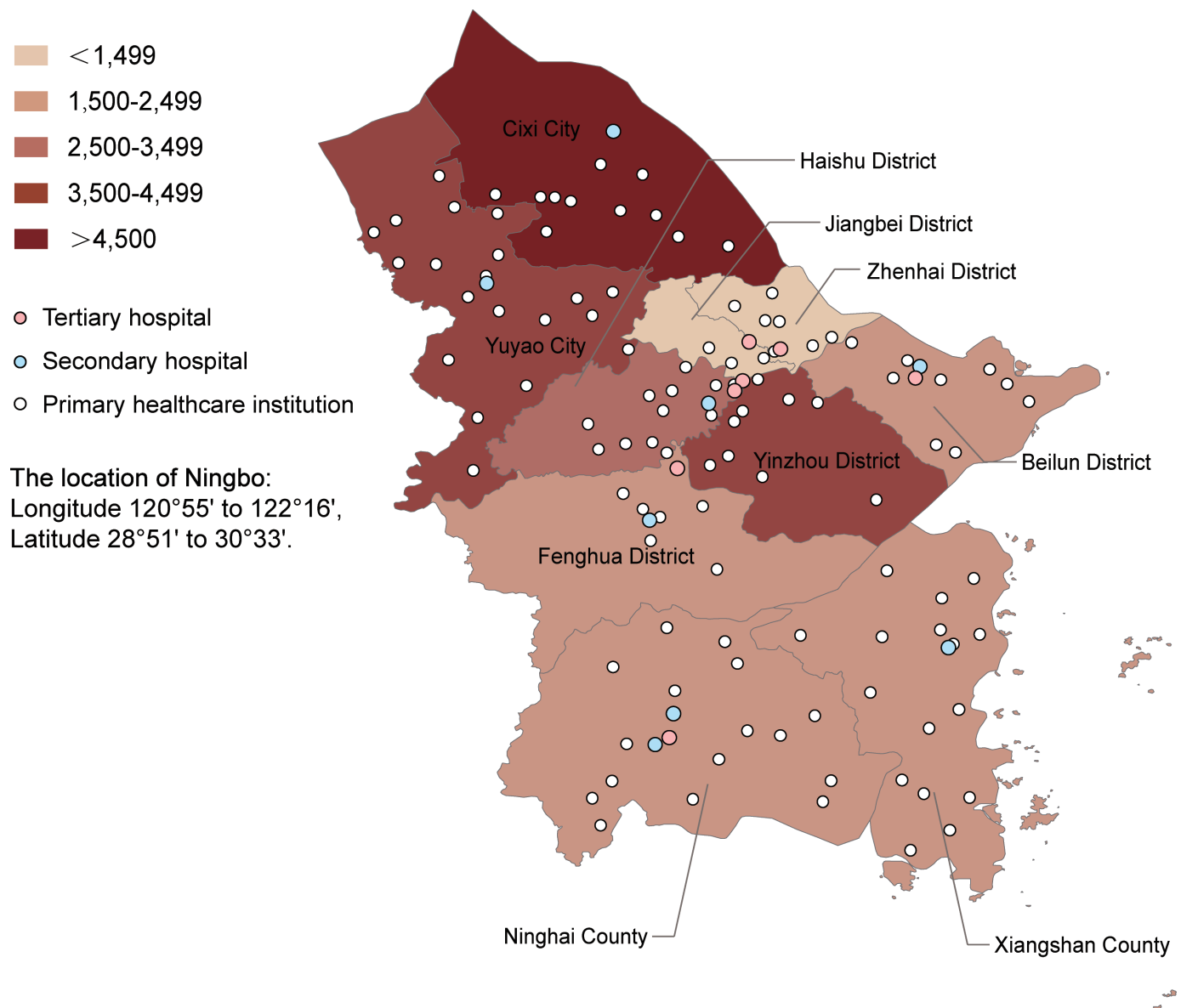
diverse array of providers and a sufficient number of qualified personnel, as shown in [figure 1](#). The city boasts nearly 2000 allied primary mental healthcare providers. The regional distribution of primary mental healthcare providers in Ningbo is presented in online supplemental SFigure 1. (2) Ningbo has a highly developed healthcare information platform. This includes the (a) NRHIP, which ranked sixth in the National Health Information Technology Development Index in 2021,<sup>24</sup> with a capacity to store 35 billion pieces of residents' medical information, including 1042 e-forms. It realises region-wide coverage, population-wide services and full life cycle management. The platform also includes (b) NMHIS, which records nearly 40 000 patients with schizophrenia of all ages, as illustrated in [figure 1](#). The system interface specification issued by the Health Commission of Ningbo guarantees that the NMHIS and NRHIP are interconnected, forming a continuous, complete and multidimensional EHR of patients with schizophrenia and encompassing several indicators such as diagnosis, medication, laboratory tests and imaging studies (eg, echocardiography). This provides a high-quality source of data with regard to the construction of NSC.

### Study design

NSC is a population-based, ambispective cohort study that includes individuals with schizophrenia in Ningbo. The follow-up for this cohort began on 1 January 2010, to bypass the initial trial period and NRHIP integration phase, as demonstrated in [figure 2](#). Clinical coding in this study was conducted by using the WHO International Classification of Diseases 10th Edition (ICD-10).<sup>25</sup> Once an individual enters the NSC, they are followed up continuously until death or relocation.

Patients were eligible if they were aged 18–65 years at diagnosis and had a diagnosis of schizophrenia confirmed by at least two independent visit records in the EHR. Additionally, they were required to provide signed informed consent, either personally or through their legal representatives, and possess complete EHR data, including date of birth, gender and a valid medical identifier. Subjects meeting any of the following criteria were not eligible to participate: (1) history of severe neurological diseases, including epilepsy, encephalitis or brain tumours or (2) history of severe head trauma resulting in unconsciousness, to exclude trauma-induced psychotic symptoms. The exclusion criteria with corresponding ICD-10 code details are provided in online supplemental S table 1.

Participants underwent a semi-structured interview to collect data on sociodemographic characteristics, comorbid mental disorders, previous psychiatric treatments, family history of psychiatric illness and childhood trauma at baseline. Schizophrenia diagnoses were confirmed using the Structured Clinical Interview for DSM-5 Disorders.<sup>26</sup> Psychopathological assessment included the Positive and Negative Syndrome Scale and Brief Psychiatric Rating Scale to evaluate the severity of psychotic symptoms.<sup>27 28</sup> The Clinical Global Impression



**Figure 1** Regional distribution of patients with schizophrenia in Ningbo and the location of mental health service organisations.

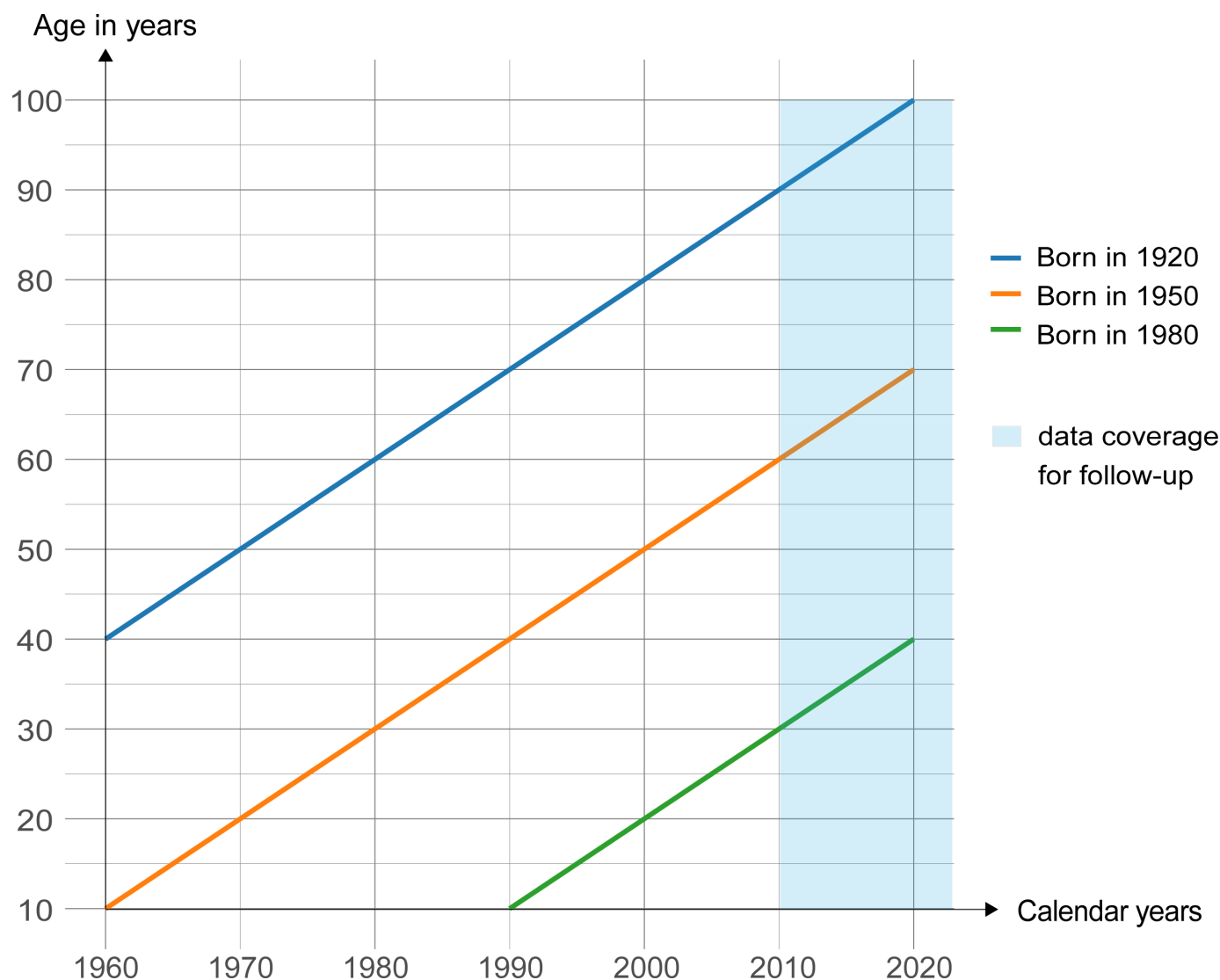
was used to assess overall disease severity,<sup>29</sup> while the Global Assessment Scale measured overall functional status.<sup>30</sup> Additionally, the Social Disability Screening Schedule was used to assess social functioning deficits in individuals with psychiatric disorders.<sup>31</sup> Risk assessment includes a history of violent, assaultive and aggressive behaviours, while a history of suicide attempts is recorded separately. If the patient is hospitalised for a long period, the Nurses' Observation Scale for Inpatient Evaluation is used to assess their behaviour and social functioning in a psychiatric inpatient setting.<sup>32</sup> Recovery is typically defined as an improvement that encompasses both the symptom and functional levels, sustained over an extended period, usually at least 2 years.<sup>33 34</sup>

#### Data sources

NSC data are primarily derived from the NMHIS, as patients with schizophrenia are enrolled in this system

by clinicians during their first visit and assigned a unique identifier. This identifier links information across multiple forms in the NMHIS, such as the patient information, patient specific file, and follow-up forms. This linkage enables access to key information on mental health variables necessary for the NSC, such as sociodemographic data, pre-existing mental disorders, risk factors for schizophrenia, psychological status, and symptoms and severity.

Furthermore, the NMHIS is linked to the NRHIP. By 31 December 2023, 97.83% of patients with schizophrenia with unique identifiers had at least one electronic medical record in the NRHIP, indicating that they had received at least one clinical service at either a hospital or a primary healthcare institution. The NRHIP supplements this information with additional key variables, such as outpatient or inpatient records and medication. The NSC leverages the strengths of both the NRHIP and NMHIS



**Figure 2** Lexis diagram of Ningbo schizophrenia cohort follow-up coverage time.

platforms, using data from the full life cycle of a patient—from pre-onset to diagnosis, follow-up, death or recovery. This comprehensive approach facilitates effective full life cycle management of patients, as illustrated in [figure 3](#).

### FINDINGS TO DATE

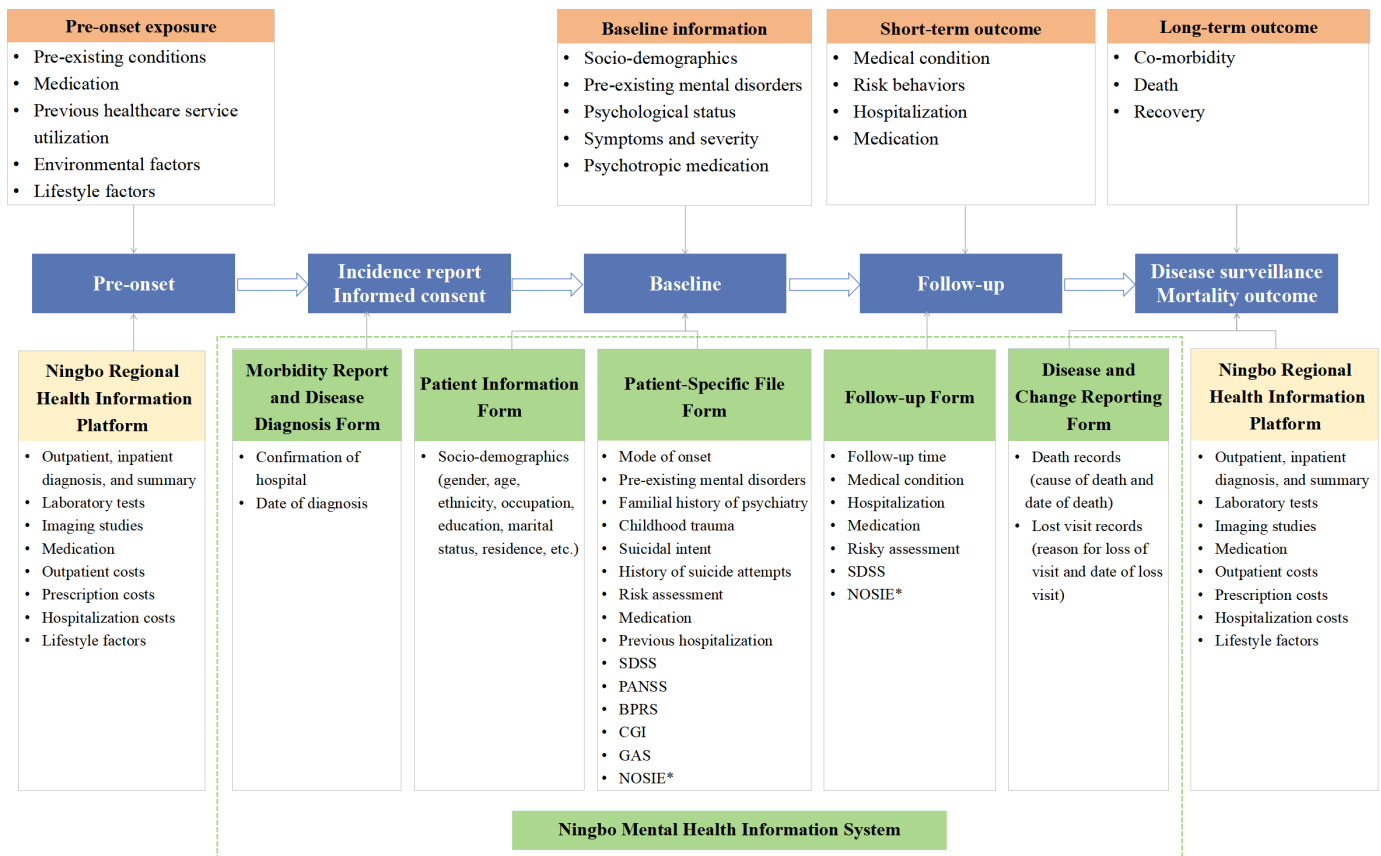
A total of 26 899 patients with schizophrenia were included in the NSC. As of 31 December 2023, there were 4505 deaths in the NSC. The median age at diagnosis was 35 years, and the age at diagnosis of the non-survivors (median (IQR): 40 (29–51) years) was significantly higher than that of the survivors (median (IQR): 34 (26–45) years). The prevalence of the disease was high in females, while their years of education were generally low; the occupational distribution was dominated by farmers and those who were laid off or unemployed, accounting for 70.5%; and unmarried, widowed and divorced patients accounted for 39.5%. There were 1503 patients with a family history, accounting for 5.6% and the mode of

onset of the disease was most acute. A detailed description is given in [table 1](#).

### LONG-TERM FOLLOW-UP

The NSC follow-up adopts a combination of active and passive mechanisms through face-to-face and outpatient follow-ups. After patients sign the informed consent form, the NMHIS sends patients' information to the primary healthcare institutions in their places of residence. The primary healthcare institutions then provide the patients with long-term follow-up and rehabilitation guidance. The initial face-to-face follow-up visit, during which informed consent is obtained, occurs within 10 working days. Each follow-up visit updates information on the measured factors, resulting in abundant longitudinal data. Meanwhile, the NRHIP will continue to collect information from patients' EHR. The dynamic measurements of key variables during follow-up in the NSC group are presented in [table 2](#).





**Figure 3** Full life-cycle management of patients in the Ningbo Schizophrenia Cohort from pre-onset to death. \*Variables were measured at hospitalisation. BRPS, Brief Psychiatric Rating Scale; CGI, Clinical Global Impression; GAS, Global Assessment Scale; NOSIE, Nurses' Observation Scale for Inpatient Evaluation; PANSS, Positive and Negative Syndrome Scale; SDSS, Social Disability Screening Schedule.

## DATA MANAGEMENT

Quality control covers the entire data collection, processing and usage processes. At the data-entry stage, a module for collecting information on patients with schizophrenia is embedded in the hospital information system of the medical institution so that the subsequent step in diagnostic and treatment activities cannot be conducted without filling in the information. This ensures that each patient's information is collected. The NMHIS is equipped with logical checking functions to ensure data validity, standardise data and reduce processing work, including data type conversion. During the data-processing phase, multiple identification codes are stored in the NSC, including the civil identification number, case number, outpatient number, inpatient number and unique identifier code used internally in the NMHIS. Links within the NMHIS forms are primarily through unique identifier codes, while links to other platforms are primarily through civil identification numbers. If the civil identification number is missing and cannot be linked, an approximate string match is conducted using other information, such as the date of birth, address, outpatient number and inpatient number. Wonder Information Corporation maintains and manages the data to ensure their integrity and availability. All personal information

of NSC patients is anonymised during the data-utilisation period.

## DATA UTILISATION

### Seeking new real-world evidence to infer the aetiology of schizophrenia

In cohort studies wherein exposure precedes and outcomes follow, the temporal sequence of causality between the exposure and outcome is an important criterion for determining causality. The NSC provides information on factors such as health service utilisation and environmental exposure across the patient's full life cycle, including pre-onset; therefore, the multidimensional pre-onset environmental and health service utilisation information obtained by the NSC may provide new real-world evidence for aetiological inferences of schizophrenia.

### Examining the comorbidity of schizophrenia to assist with managing comorbidities

The comorbidity rate of individuals with mental disorders exceeds 35%<sup>35</sup> as a result of poor access to health services and the side effects of antipsychotic drugs. Single-disease management cannot minimise disruptive medicine, which wastes medical resources and is not conducive to

**Table 1** Baseline characteristics of the Ningbo Schizophrenia Cohort

	Survivors	Non-survivors	Overall
	N=22 394	N=4505	N=26 899
	N (%)	N (%)	N (%)
Sex			
Male	9805 (43.8)	2184 (48.5)	11 989 (44.6)
Female	12 589 (56.2)	2321 (51.5)	14 910 (55.4)
Age at diagnosis (years)			
<29	8012 (35.8)	1188 (26.4)	9200 (34.2)
30–39	6037 (27.0)	1020 (22.6)	7057 (26.2)
40–49	4450 (19.9)	1019 (22.6)	5469 (20.3)
>50	3895 (17.4)	1278 (28.4)	5173 (19.2)
Median (IQR)	34 (26–45)	40 (29–51)	35 (26–46)
Marital status			
Unmarried	5699 (25.5)	1005 (22.3)	6704 (24.9)
Married	13 654 (61.0)	2582 (57.3)	16 236 (60.4)
Widowed	1005 (4.5)	570 (12.7)	1575 (5.9)
Divorced	2015 (9.0)	335 (7.4)	2350 (8.7)
Unclear	21 (0.1)	13 (0.3)	34 (0.1)
Years of education			
<7	10 925 (48.8)	3352 (74.4)	14 277 (53.1)
7–9	7094 (31.7)	815 (18.1)	7909 (29.4)
10–12	2406 (10.7)	206 (4.6)	2612 (9.7)
>12	1337 (6.0)	67 (1.5)	1404 (5.2)
Unclear	632 (2.8)	65 (1.4)	697 (2.6)
Ethnicity			
Han	22 350 (99.8)	4502 (99.9)	26 852 (99.8)
Others	44 (0.2)	3 (0.1)	47 (0.2)
Occupation			
Workers	2131 (9.5)	125 (2.8)	2256 (8.4)
Managers	33 (0.1)	1 (0.0)	34 (0.1)
Farmers	8868 (39.6)	2542 (56.4)	11 410 (42.4)
Laid-off or unemployed	6567 (29.3)	1000 (22.2)	7567 (28.1)
Student	35 (0.2)	3 (0.1)	38 (0.1)
Retired	1617 (7.2)	297 (6.6)	1914 (7.1)
Professionals and technicians	184 (0.8)	14 (0.3)	198 (0.7)
Others	1969 (8.8)	453 (10.1)	2422 (9.0)
Unclear	990 (4.4)	70 (1.6)	1060 (3.9)
Economic status			
Poor, below local poverty line	5515 (24.6)	1559 (34.6)	7074 (26.3)
Non-poor	16 879 (75.4)	2915 (64.7)	19 794 (73.6)
Unclear	0 (0.0)	31 (0.7)	31 (0.1)
Familial history of psychiatry			
No	20 331 (90.8)	4158 (92.3)	24 489 (91.0)
Yes	1249 (5.6)	254 (5.6)	1503 (5.6)
Unclear	814 (3.6)	93 (2.1)	907 (3.4)
Mode of onset			

Continued

**Table 1** Continued

	Survivors	Non-survivors	Overall
	N=22 394	N=4505	N=26 899
	N (%)	N (%)	N (%)
Acute	9046 (40.4)	2578 (57.2)	11 624 (43.2)
Subacute	6208 (27.7)	892 (19.8)	7100 (26.4)
Chronic	7138 (31.9)	1003 (22.3)	8141 (30.3)
Unclear	2 (0.0)	32 (0.7)	34 (0.1)
Guardianship			
No	2264 (10.1)	496 (11.0)	2760 (10.3)
Yes	20 130 (89.9)	4009 (89.0)	24 139 (89.7)

patient rehabilitation; therefore, comorbidity management is imperative. The NSC provides the times of initial and comorbidity diagnoses. Therefore, the NSC may provide real-world data for the identification of comorbidity clusters and the analysis of comorbidity trajectories in patients with late-stage schizophrenia.

### Constructing a predictive model for schizophrenia and subsequent CVD

CVD is a serious comorbidity of schizophrenia, and its high prevalence is the primary reason for the reduced life expectancy of these patients.<sup>36</sup> However, existing CVD risk-prediction models, such as the Framingham Risk Score,<sup>37–39</sup> are not suitable for risk prediction in patients with schizophrenia as they do not consider specific risk predictors such as anti-psychotic drug utilisation. The NSC contains information on predictors from the Prediction for Atherosclerotic CVD Risk in China<sup>40</sup>; WHO CVD risk charts<sup>41</sup>; and specific predictors such as anti-psychotic drug utilisation, MRI and cardiac auxiliary examination. Therefore, the NSC contains a database for developing predictive models for schizophrenia and subsequent CVD.

### Identifying the best interventions available for patients with schizophrenia

Currently, the most important treatment for schizophrenia is the administration of antipsychotic medication, but medication may produce side effects and have little effect on negative symptoms.<sup>42</sup> In this scenario, psychosocial interventions (eg, family interventions, family psychoeducation and cognitive-behavioural therapies) play an important role in the rehabilitation of schizophrenia patients<sup>43</sup> and are an important means of supplementing medication. The applicable intervention methods vary due to differences in the mode of onset of the disease in patients and their economic status; therefore, we hope to identify the best intervention methods for patients with different types of schizophrenia by comparing the patient efficacy of the different methods using the data resources of the NSC.

### DISCUSSION

The NSC is an ambispective, observational, multicentre cohort study focused on patients with schizophrenia. This study covers the full life cycle of patients, accumulating rich epidemiological and clinical data. The NSC aims to provide a database for future research on schizophrenia, determine the burden of disease and risk factors, improve prognosis and reduce the risk of relapse. Notably, the median age of onset of schizophrenia is generally early adulthood,<sup>44 45</sup> whereas in the NSC, the median age of diagnosis is 34 years, as the diagnostic process for schizophrenia in Ningbo may require up to 10 years or more. This phenomenon may be influenced by a variety of factors, including the inadequacy of social support systems and social and economic factors. This is in addition to the cultural perceptions, accessibility of healthcare resources and level of public awareness of mental health issues.

### Strengths and limitations

The key strength of the NSC lies in its foundation as an EHR-based cohort, enabling data collection throughout patients' full life cycles—from pre-onset and diagnosis to follow-up, recovery or death. This design provides a broad range of study factors and outcomes. Given that mental disorders are significant NCDs, the repeated quantitative characteristics obtained from follow-up allow for the assessment of treatment effects. The longitudinal data, with their temporal sequence, enable a comprehensive evaluation of risk factors for disease development and various factors related to prognosis, facilitating aetiological research on schizophrenia.

Additionally, the NSC collects diverse, multidimensional data (eg, laboratory tests and imaging studies) by accessing electronic medical records. This valuable resource assists in identifying whether functional abnormalities of other organs, pathological lesions or metabolic abnormalities are associated with schizophrenia. Therefore, numerous variables and outcomes can be extracted while conducting a real-world study.

Large sample size and broad coverage are another major advantage of the NSC. The NMHIS is embedded in the hospital information system of all psychiatric hospitals

**Table 2** Dynamic measurement of key variables for the Ningbo Schizophrenia Cohort follow-up

Type of information	Variables	Baseline	Follow-up
Sociodemographics	Sex, age, ethnicity, occupation, education years, marital status, economic status, type of health insurance and residence	*	
Pre-existing mental disorders	Diagnosis of other mental disorders	*	
	Previous forms of psychiatric treatment	*	
Risk factors for schizophrenia	Familial psychiatry history	*	
	Childhood trauma	*	
	Socioeconomic status	*	
	Season of birth	*	
	Urbanicity	*	
	Green spaces	*	
	Air pollutants	*	
	Body mass index	*	*
Lifestyle factors	Smoking	*	
	Alcohol consumption	*	
	Physical activity	*	
Psychological status	Suicidal intent	*	
	History of suicide attempts	*	
	Positive and Negative Syndrome Scale	*	
	Brief Psychiatric Rating Scale	*	
	Duration of untreated psychosis	*	
Symptoms and severity	Age of first psychiatric hospitalisation	*	
	Mode of onset	*	
	Clinical Global Impression	*	
	Global Assessment Scale	*	
	Social Disability Screening Schedule	*	*
	Number of psychiatric outpatient visits	*	*
	Risk assessment	*	*
	Number of psychiatric admissions and length of stay	*	*
	Date of last psychiatric discharge	*	*
	Nurses' Observation Scale for Inpatient Evaluation	*†	*†
Psychotropic medication	Time to first antipsychotic medication	*	*
	Duration of medication	*	
	Drug name	*	
	Dose	*	
	Chlorpromazine equivalent		*
	Defined daily dose <sup>51</sup>		*
	Adherence to medication		*
	Adverse drug reactions		*
Death	Cause and date of death		*
Medical condition	Other comorbidities (eg, cardiovascular disease, gastrointestinal system disease, endocrine system disease and neurological system disease)	*	*
	Medication	*	*

Continued



**Table 2** Continued

Type of information	Variables	Baseline	Follow-up
Ancillary tests	Laboratory tests	*	*
	Imaging studies (eg, CT and MRI)	*	*
Health economics	Outpatient costs	*	*
	Prescription costs	*	*
	Hospitalisation costs	*	*†
*Variables were collected at this stage.			
†Variables were measured at hospitalisation.			

in Ningbo, and patients diagnosed with schizophrenia are included in the cohort after providing informed consent within the hospital setting. This makes this study less susceptible to selection bias, ensuring the accuracy and representativeness of the study cohort.

This study has some limitations. First, the NSC is a real-world research cohort with observational and non-randomised interventions, and the real-world data to be studied may not cover the entire study population. Furthermore, there is potential information bias in this study, which refers to systematic errors in data collection, recording or classification that could affect the accuracy of study findings.<sup>46</sup> Schizophrenia diagnoses in this study were confirmed using the Structured Clinical Interview for DSM-5 Disorders, ensuring diagnostic accuracy. However, medical conditions obtained from the NRHIP were coded using ICD-10, which may introduce misclassification bias in comorbid conditions due to potential coding inaccuracies in routine clinical practice. This is an inherent limitation of using EHR data for research.<sup>17</sup> ICD-10 codes can be combined with natural language processing to accurately identify patients.<sup>47</sup> Third, perinatal factors and substance use have been identified as potential risk factors for schizophrenia<sup>48–50</sup>; however, due to privacy and ethical review constraints, such information was not included in this study. Future research could explore these aspects further by integrating maternal and child health databases, implementing drug screening or conducting long-term cohort studies. Finally, the exclusion of organic diseases is crucial in schizophrenia research. In this study, we excluded only conditions that would significantly impact schizophrenia diagnoses. To ensure data completeness, we implemented a data flagging strategy for patients newly diagnosed with neurological disorders (eg, Niemann-Pick disease, Huntington's disease, Wilson's disease, Dandy-Walker syndrome and Alzheimer's disease) during follow-up. These cases are flagged in the NSC, allowing future researchers to determine their inclusion based on specific study objectives.

**Acknowledgements** The researchers thank all the patients who participated in the NSC.

**Contributors** He Ye contributed significantly to the study design and drafted the manuscript; Yang Zhao contributed to the study conceptualisation and design; Lian Li, Yi Qian and Hangjie Zhu contributed to data collection and collation; Guolin Bian and Liya Liu contributed significantly to the conceptualisation of the study, and both

contributed equally to the guidance and supervision of this research. All authors revised the manuscript. Liya Liu serve as the guarantor of this study, ensuring the accuracy and integrity of the work.

**Funding** This study was funded by the National Natural Science Foundation of China (82373690, 81602940), National Key Research and Development Programme of China (2022YFC3702702), Natural Science Foundation of Zhejiang Province (ZCLY24H2601), One Health Interdisciplinary Research Project, Ningbo University, State Key Laboratory of Reproductive Medicine, Nanjing Medical University (SKLRMK202104), Key Laboratory of Modern Toxicology of Ministry of Education, Nanjing Medical University and Ningbo Medical & Health Leading Academic Discipline Project (2022-F28).

**Map disclaimer** The depiction of boundaries on this map does not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. This map is provided without any warranty of any kind, either express or implied.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Ethics approval** This study involves human participants and was approved by the Medical Ethics Committee of Ningbo University (NBU-2024-011) and registered at ClinicalTrials.gov (NCT06370793). All patients signed an informed consent form of their own accord by themselves or through their guardians prior to being included in the cohort. Access to information was based on a specially constructed server, which could not be accessed without approval, and confidentiality agreements were signed by the researchers prior to accessing the data.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data may be obtained from a third party and are not publicly available. The NSC was developed by the Affiliated Kangning Hospital of Ningbo University, which was responsible for granting access to and using patient data. Access to data resources requires completion of a proposal that includes the proposed research programme and the primary source of funding. All research must be ethically approved, and the feasibility as well as compliance of the research programme with the data-access policy must be assessed. Interested researchers are encouraged to contact the principal investigator, Professor Liya Liu (liuliya@nbu.edu.cn).

**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 Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iD

Liya Liu <http://orcid.org/0000-0001-9105-9603>

## REFERENCES

- Institute of health Metrics and Evaluation. Ment health (lond). 2021. Available: <https://www.healthdata.org/research-analysis/health-risks-issues/mental-health>
- GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry* 2022;9:137–50.
- 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.
- Chisholm D, World Health Organization, Geneva, Switzerland. Investing in mental health. *Easter Mediterr Health J* 2015;12:531–4.
- Patel V, Chisholm D, Parikh R, et al. Addressing the burden of mental, neurological, and substance use disorders: key messages from Disease Control Priorities, 3rd edition. *The Lancet* 2016;387:1672–85.
- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5*. 5th edn. Arlington, VA, US: American Psychiatric Publishing, Inc, 2013:90–1.
- Institute for Health Metrics and Evaluation. Schizophrenia—level 3 cause 2021. 2021. Available: <https://www.healthdata.org/research-analysis/diseases-injuries-risks/factsheets/2021-schizophrenia-level-3-disease>
- Huang Y, Wang Y, Wang H, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. *Lancet Psychiatry* 2019;6:211–24.
- Brown S, Kim M, Mitchell C, et al. Twenty-five year mortality of a community cohort with schizophrenia. *Br J Psychiatry* 2010;196:116–21.
- Lieberman JA, Small SA, Girgis RR. Early Detection and Preventive Intervention in Schizophrenia: From Fantasy to Reality. *Am J Psychiatry* 2019;176:794–810.
- Tsao CW, Vasan RS. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *Int J Epidemiol* 2015;44:1800–13.
- FitzGerald GA. Evolution in translational science: Whither the CTSAs? *Sci Transl Med* 2015;7:284fs15.
- Sun X, Tan J, Tang L, et al. Real world evidence: experience and lessons from China. *BMJ* 2018;360:j5262.
- Vassy JL, Ho Y-L, Honerlaw J, et al. Yield and bias in defining a cohort study baseline from electronic health record data. *J Biomed Inform* 2018;78:54–9.
- Petrazzini BO, Chaudhary K, Márquez-Luna C, et al. Coronary Risk Estimation Based on Clinical Data in Electronic Health Records. *J Am Coll Cardiol* 2022;79:1155–66.
- Wang J, Bao B, Shen P, et al. Using electronic health record data to establish a chronic kidney disease surveillance system in China: protocol for the China Kidney Disease Network (CK-NET)-Yinzhou Study. *BMJ Open* 2019;9:e030102.
- Scierka LE, Bradley BA, Glynn E, et al. Chronic Cough: Characterizing and Quantifying Burden in Adults Using a Nationwide Electronic Health Records Database. *J Healthc Inform Res* 2024;8:50–64.
- Abrahão MTF, Nobre MRC, Gutierrez MA. A method for cohort selection of cardiovascular disease records from an electronic health record system. *Int J Med Inform* 2017;102:138–49.
- Lin H, Tang X, Shen P, et al. Using big data to improve cardiovascular care and outcomes in China: a protocol for the Chinese Electronic health Records Research in Yinzhou (CHERRY) Study. *BMJ Open* 2018;8:e019698.
- Garriga R, Mas J, Abraha S, et al. Machine learning model to predict mental health crises from electronic health records. *Nat Med* 2022;28:1240–8.
- DeFerio J, Banerjee S, Alexopoulos G, et al. Using Electronic Health Records and Machine Learning to Predict Incident Psychiatric Hospitalization. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2020;87:S68–9.
- Ningbo Municipal Bureau of statistics. Yangtze River Delta Cities Data 2024. Available: [http://tjj.ningbo.gov.cn/art/2025/2/18/art\\_1229042823\\_58920984.html](http://tjj.ningbo.gov.cn/art/2025/2/18/art_1229042823_58920984.html)
- Ningbo Municipal Bureau of statistics. Ningbo Population Key Data Bulletin 2024. Available: [http://tjj.ningbo.gov.cn/art/2025/3/10/art\\_1229042825\\_58921023.html](http://tjj.ningbo.gov.cn/art/2025/3/10/art_1229042825_58921023.html)
- National Health Commission of the People's Republic of China. National health and health informatisation development index city results released. 2021. Available: <http://www.nhc.gov.cn/cms-search/xxgk/getManuscriptXxgk.htm?id=1d7149a63b784b04b6aaa8052b6b30e5>
- World Health Organization. International classification of diseases. 2019. Available: <https://icd.who.int/browse10/2019/en#/F70-F79>
- First MB. Structured clinical interview for the (SCID). In: *The Encyclopedia of Clinical Psychology*. 2015: 1–6.
- Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* 1987;13:261–76.
- Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. *Psychol Rep* 1962;10:799–812.
- William G. *ECDEU assessment manual for psychopharmacology*. Rockville, MD: US Department of Health, Education, and Welfare, 1976:33:766–71.
- Endicott J, Spitzer RL, Fleiss JL, et al. The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. *Arch Gen Psychiatry* 1976;33:766–71.
- Zhang M, He Y. *Manual of the psychiatric rating scale*. 2015: 163–6.
- Lyall D, Hawley C, Scott K. Nurses' Observation Scale for Inpatient Evaluation: reliability update. *J Adv Nurs* 2004;46:390–4.
- Jääskeläinen E, Juola P, Hirvonen N, et al. A systematic review and meta-analysis of recovery in schizophrenia. *Schizophr Bull* 2013;39:1296–306.
- Lieberman RP, Kopelowicz A, Ventura J, et al. Operational criteria and factors related to recovery from schizophrenia. *Int Rev Psychiatry* 2002;14:256–72.
- Barnett K, Mercer SW, Norbury M, et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* 2012;380:37–43.
- Lambert AM, Parretti HM, Pearce E, et al. Temporal trends in associations between severe mental illness and risk of cardiovascular disease: A systematic review and meta-analysis. *PLoS Med* 2022;19:e1003960.
- Liu J, Hong Y, D'Agostino RB Sr, et al. Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-Provincial Cohort Study. *JAMA* 2004;291:2591–9.
- Wilson PW, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97:1837–47.
- Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: The Framingham study. *Am J Cardiol* 1976;38:46–51.
- Yang X, Li J, Hu D, et al. Predicting the 10-Year Risks of Atherosclerotic Cardiovascular Disease in Chinese Population: The China-PAR Project (Prediction for ASCVD Risk in China). *Circulation* 2016;134:1430–40.
- Kaptoge S, Pennells L, De Bacquer D, et al. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *Lancet Glob Health* 2019;7:e1332–45.
- Stępnicki P, Kondej M, Kaczor AA. Current Concepts and Treatments of Schizophrenia. *Molecules* 2018;23:2087.
- Ashcroft K, Kim E, Elephant E, et al. Meta-Analysis of Caregiver-Directed Psychosocial Interventions for Schizophrenia. *Community Ment Health J* 2018;54:983–91.
- McGrath JJ, Al-Hamzawi A, Alonso J, et al. Age of onset and cumulative risk of mental disorders: a cross-national analysis of population surveys from 29 countries. *Lancet Psychiatry* 2023;10:668–81.
- Grande I, Berk M, Birmaher B, et al. Bipolar disorder. *Lancet* 2016;387:1561–72.
- Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002;359:248–52.
- Weiner M, Dexter PR, Heithoff K, et al. Identifying and Characterizing a Chronic Cough Cohort Through Electronic Health Records. *Chest* 2021;159:2346–55.
- Khandaker GM, Zimbron J, Dalman C, et al. Childhood infection and adult schizophrenia: a meta-analysis of population-based studies. *Schizophr Res* 2012;139:161–8.
- Mueser KT, McGurk SR. Schizophrenia. *The Lancet* 2004;363:2063–72.
- Crawford P, Go KV. Schizophrenia. *Am Fam Physician* 2022;106:388–96.
- Leucht S, Samara M, Heres S, et al. Dose Equivalents for Antipsychotic Drugs: The DDD Method: Table 1. *SCHBUL* 2016;42:S90–4.