

Original Paper

Estimating the Prevalence of Schizophrenia in the General Population of Japan Using an Artificial Neural Network–Based Schizophrenia Classifier: Web-Based Cross-Sectional Survey

Pichsinee Choomung¹, PharmD; Yupeng He¹, MD, PhD; Masaaki Matsunaga¹, MD, PhD; Kenji Sakuma², MD, PhD; Taro Kishi², MD, PhD; Yuanying Li³, PhD; Shinichi Tanihara⁴, PhD; Nakao Iwata², MD, PhD; Atsuhiko Ota¹, MD, PhD

¹Department of Public Health, Fujita Health University School of Medicine, Toyoake, Japan

²Department of Psychiatry, Fujita Health University School of Medicine, Toyoake, Japan

³Department of Public Health and Health Systems, Nagoya University Graduate School of Medicine, Nagoya, Japan

⁴Department of Public Health, Kurume University School of Medicine, Kurume, Japan

Corresponding Author:

Yupeng He, MD, PhD
Department of Public Health
Fujita Health University School of Medicine
1-98 Dengakugakubo, Kutsukake-cho
Toyoake, 470-1192
Japan
Phone: 81 562-93-2476
Fax: 81 562-93-3079
Email: yupeng.he@fujita-hu.ac.jp

Abstract

Background: Estimating the prevalence of schizophrenia in the general population remains a challenge worldwide, as well as in Japan. Few studies have estimated schizophrenia prevalence in the Japanese population and have often relied on reports from hospitals and self-reported physician diagnoses or typical schizophrenia symptoms. These approaches are likely to underestimate the true prevalence owing to stigma, poor insight, or lack of access to health care among respondents. To address these issues, we previously developed an artificial neural network (ANN)–based schizophrenia classification model (SZ classifier) using data from a large-scale Japanese web-based survey to enhance the comprehensiveness of schizophrenia case identification in the general population. In addition, we also plan to introduce a population-based survey to collect general information and sample participants matching the population’s demographic structure, thereby achieving a precise estimate of the prevalence of schizophrenia in Japan.

Objective: This study aimed to estimate the prevalence of schizophrenia by applying the SZ classifier to random samples from the Japanese population.

Methods: We randomly selected a sample of 750 participants where the age, sex, and regional distributions were similar to Japan’s demographic structure from a large-scale Japanese web-based survey. Demographic data, health-related backgrounds, physical comorbidities, psychiatric comorbidities, and social comorbidities were collected and applied to the SZ classifier, as this information was also used for developing the SZ classifier. The crude prevalence of schizophrenia was calculated through the proportion of positive cases detected by the SZ classifier. The crude estimate was further refined by excluding false-positive cases and including false-negative cases to determine the actual prevalence of schizophrenia.

Results: Out of 750 participants, 62 were classified as schizophrenia cases by the SZ classifier, resulting in a crude prevalence of schizophrenia in the general population of Japan of 8.3% (95% CI 6.6%-10.1%). Among these 62 cases, 53 were presumed to be false positives, and 3 were presumed to be false negatives. After adjustment, the actual prevalence of schizophrenia in the general population was estimated to be 1.6% (95% CI 0.7%-2.5%).

Conclusions: This estimated prevalence was slightly higher than that reported in previous studies, possibly due to a more comprehensive disease classification methodology or, conversely, model limitations. This study demonstrates the capability

of an ANN-based model to improve the estimation of schizophrenia prevalence in the general population, offering a novel approach to public health analysis.

JMIR Form Res 2025;9:e66330; doi: [10.2196/66330](https://doi.org/10.2196/66330)

Keywords: schizophrenia; schizophrenic; prevalence; artificial neural network; neural network; neural networks; ANN; deep learning; machine learning; SZ classifier; web-based survey; epidemiology; epidemiological; Japan; classifiers; mental illness; mental disorder; mental health

Introduction

Schizophrenia is a complex psychiatric disorder characterized by significant impairments in the perception of reality and behavioral changes. It includes positive (eg, delusions, hallucinations, disorganized thinking, and notably disorganized behaviors) and negative symptoms (eg, limited speech, decreased emotional expression, inability to experience interest or pleasure, and social withdrawal). Affected individuals often experience cognitive impairments affecting memory, attention, and problem-solving abilities [1,2]. The Global Burden of Disease Study [3] reported that schizophrenia accounted for 23.6 million estimated cases in 2019, with an age-standardized prevalence rate of 287.4 per 100,000 people. Median estimates from internationally published studies indicate a 12-month prevalence of 0.33%, lifetime prevalence of 0.48%, and point prevalence of 0.32% [4]. However, there is a lack of published studies on the prevalence of schizophrenia in the general Japanese population. Only a few population-based studies have estimated the prevalence of schizophrenia in Japan. For instance, the point prevalence of schizophrenia stratified by sex and age ranges from 0.4% to 1.1% [5]. However, these figures were derived from the Vital Statistics and the Patient Survey in 2008, which probably do not represent the general population since the Patient Survey could not encompass patients who do not visit a doctor. The estimated lifetime prevalence of schizophrenia is 0.59%, according to a secondary analysis of the 2019 cross-sectional Japan National Health and Wellness Survey [6]. However, this study may produce underestimation since it only depended on physician diagnoses. Individuals with schizophrenia hardly self-identify as having schizophrenia because of stigma and lack of awareness of schizophrenia and its symptoms [7-13].

In our previous research, we developed an artificial neural network (ANN)-based schizophrenia classification model (SZ classifier) to classify schizophrenia cases in the Japanese population and verified its generalizability [14]. The model was trained by using data from a large-scale Japanese web-based survey; the presence of schizophrenia served as a response variable. The general information including demographics, health-related background, physical comorbidities, psychiatric comorbidities, and social comorbidities were collected and used as feature variables. This SZ classifier achieved high internal validity with a high area under the receiver operating characteristic curve (AUC) of 0.86 [14]. In another study, we also assessed the SZ classifier's external validity by applying it to psychiatric outpatients, achieving a sensitivity of 75% [15]. The SZ classifier suggests a potential

for classifying schizophrenia cases in the general Japanese population.

In this study, we aimed to estimate the prevalence of schizophrenia in the general Japanese population using the SZ classifier through the following three approaches: (1) conducting a population-based survey to recruit participants and collect their general information, (2) sampling participants who match the distribution to the population demographic structure, and (3) using a pretrained machine learning model to detect schizophrenia cases. Study participants were randomly sampled from a pooled panel. The samples are matched to the demographic distribution of the Japanese population. General information, the same as that for the development of the SZ classifier, was collected through the web-based survey for each study participant. The estimation of the schizophrenia prevalence was implemented through the following two steps: (1) calculating the crude prevalence based on crude cases that are detected by the SZ classifier and (2) refining the crude prevalence by excluding false-positive estimates and including the false-negative estimates.

Methods

Development of the SZ Classifier

The SZ classifier was developed as part of a prior study [14] using a dataset obtained from an internet research agency's pooled panels (Rakuten Insight, Inc) [16]. The dataset included 223 individuals with schizophrenia and 1776 healthy controls. For each sample, schizophrenia status served as the response variable, while demographic details, health-related background information, and physical, psychiatric, and social comorbidities were considered feature variables. Evidence on the association between these feature variables and schizophrenia has been previously published [17,18]. For example, demographic and health-related characteristics included age, BMI, smoking status, restriction in functional capacity, and bowel frequency. Physical comorbidities included overweight, cancer, cardiovascular disease, heart failure, and hypertension. Psychiatric comorbidities included frequency of sleep medication use, bedtime, hours of sleep, perceived stress, and positive reasons for living. The social comorbidities included educational background, type of occupation, type of employment, household income, and marital status. The inclusion criteria of the study participants and definitions of the feature variables are detailed in previous publications [14,18] and provided in [Multimedia Appendix 1](#).

The SZ classifier was constructed using an ANN concept. This multilayer perceptron neural network model comprised 5 hidden layers with 128, 64, 32, 16, and 8 neurons in each respective layer. It used the HeNormal weight initializer and the ReLU activation function in the hidden layers. A sigmoid activation function was used in the output layer. Comprehensive details can be found in a prior publication [14]. The model's performance was evaluated using the AUC score. During internal validation, the SZ classifier achieved an AUC of 0.86, accuracy of 0.93, specificity of 0.96, and sensitivity of 0.56. He et al [14] performed 10,000 bootstrapping iterations and a 10-fold cross-validation to further evaluate the algorithm's performance and achieved AUCs of 0.847 and 0.81, respectively.

External Validation of the SZ Classifier

External validation of the SZ classifier was conducted by applying it to a sample of patients who visited the psychiatric department of Fujita Health University Hospital between January 2022 and May 2023 [15]. Patients aged 20-75 years diagnosed with schizophrenia, major depressive disorder (MDD), bipolar disorder (BD), or obsessive-compulsive disorder (OCD) were selected. Therefore, 150 participants were included in the study: 61 with schizophrenia, 56 with MDD, 32 with BD, and one with OCD. The participants were required to complete the same survey used to develop the SZ classifier to predict the likelihood of schizophrenia. The external validation process revealed that the SZ classifier achieved a sensitivity of 0.75 but misclassified 55.4% of patients with MDD and 59.4% of patients with BD as having schizophrenia [15].

Estimation of the Prevalence of Schizophrenia by Using the SZ Classifier

A cross-sectional study was conducted using data from a large-scale web-based survey in Japan. The number of participants was set to 750. The sample size was decided using the single population proportion formula ($n = [Z^2 \cdot p(1-p)] / e^2$) by assuming a 95% CI ($Z=1.96$), 1% margin of error ($e=0.01$), and 2% proportion of schizophrenia ($p=2%$) based on previous studies. Participants were randomly recruited from those registered as survey monitors with

Rakuten Insight, a large-scale web-based survey incorporated approximately 2.2 million registered members as of September 2022 and provided reliable data with regular quality maintenance [16]. Specifically, the recruited criteria were set to the combination of 3 items of age intervals (20-75 years old, divided into 5 distinct age groups), gender (male and female), and regions (13 geographical administrative regions in Japan: Hokkaido, Tohoku, North Kanto, Greater Tokyo Area, Koshinetsu, Hokuriku, Tokai, Kinki excluding Keihanshin, Keihanshin, Chugoku, Shikoku, Kyushu excluding Okinawa, and Okinawa). For each combination category of these 3 items, the number of participants was set to match the composition of the national population as of October 1, 2021, as estimated by the Ministry of Internal Affairs and Communications, Japan. Participants were recruited until the target number for each combination category was reached. The recruitment period was from November 20 to 26, 2023. This was a web-based survey in which only authorized panelists were allowed to respond to the survey to prevent fraudulent registrants from participating. To administer the survey, respondents had to log in to their accounts to prevent multiple entries from the same individual. They were required to answer 38 questions (some questions had several subquestions) that are distributed into 38 pages. Respondents were asked to review their responses and ensure completeness before submission. Specifically, the web-based survey system notified respondents if there were any blank answers. The participation rate for this survey was estimated to be 0.30, and the completion rate was not calculated because if the respondents terminated during the answering process (ie, logged out, closed the web page, etc), they were not counted in the statistics. Incomplete questionnaires or questionnaires that were submitted too soon were not used.

Each participant's demographic information, health-related backgrounds, and physical, psychiatric, and social comorbidities were collected and applied to the SZ classifier, yielding the predicted crude number of schizophrenia cases and crude prevalence of schizophrenia. [Textbox 1](#) shows a summary of the feature variable categories collected from study participants. These variables are the same as those used in developing the SZ classifier. For detailed information on these variables, please refer to [Multimedia Appendix 1](#).

Textbox 1. Summary of categories of feature variables collected from study participants, including demographic, health-related backgrounds, physical comorbidities, psychiatric comorbidities, and social comorbidities.

Demographic, health-related backgrounds, and physical comorbidities:

- Age
- BMI
- Smoking
- Alcohol drinking
- Sports
- Eating behaviors
- Bowel movement
- Restriction in functional capacity
- Self-rated health status
- Physical comorbidities

Psychiatric comorbidities:

- Frequency of sleep medication use
- Bedtime
- Sleep hours and qualities
- Depressive symptoms
- Perceived stress
- Positive reason for living (ikigai)
- Internet use hours per week

Social comorbidities:

- Regularity of health checkups
- Educational background
- Household income
- Type of employment and occupation
- Marital status and family structure
- Social support and social capital

We further refined these crude estimates by excluding false positives and adding false negatives. This was based on the model's internal and external validation results and literature reviews. Details of the model's internal and external validation results are provided in the [Multimedia Appendix 1](#).

False-positive cases included individuals without mental disorders who were misclassified as having schizophrenia and those with other mental disorders who were misclassified as having schizophrenia. The false-positive cases were calculated as follows:

$$N_{fp} = (P_{hc} \cdot r_{hc} + P_{MDD} \cdot r_{MDD} + P_{BD} \cdot r_{BD}) \cdot n$$

where N_{fp} represents the number of false-positive cases, P_{hc} represents the percentage of individuals who have not had any mental disorders in the population, r_{hc} represents the rate at which the SZ classifier falsely classified individuals without mental disorders as having schizophrenia, P_{MDD} represents the prevalence of MDD in the population, r_{MDD} represents the rate at which the SZ classifier falsely classified individuals with MDD as having schizophrenia, P_{BD} represents the prevalence of BD in the population, r_{BD} represents the rate at which the SZ classifier falsely classified individuals with BD as having schizophrenia, n represents the sample size, which included 750 participants.

The SZ classifier falsely classified 3.1% of the individuals without mental disorders as having schizophrenia [14]. The SZ classifier misclassified 55.4% of the patients with MDD and 59.4% of the patients with BD as having schizophrenia

[15]. According to a literature review, 80% of the general population in Japan does not have mental disorders [19]. The point prevalence of MDD and BD was 7.9% [20] and 0.4% [21], respectively; namely, $P_{hc}=80\%$, $r_{hc}=3.1\%$, $P_{MDD}=7.9\%$, $r_{MDD}=55.4\%$, $P_{BD}=0.4\%$, $r_{BD}=59.4\%$, and $n=750$.

False-negative cases included individuals with schizophrenia but were not classified as schizophrenia cases. The proportion of true-positive to false-negative cases in the external validation was 46:15 [15]. Hence, the false-negative cases can be calculated as follows:

$$N_{tp} = N_{crude} - N_{fn}$$

$$N_{tp}/N_{fn} = p$$

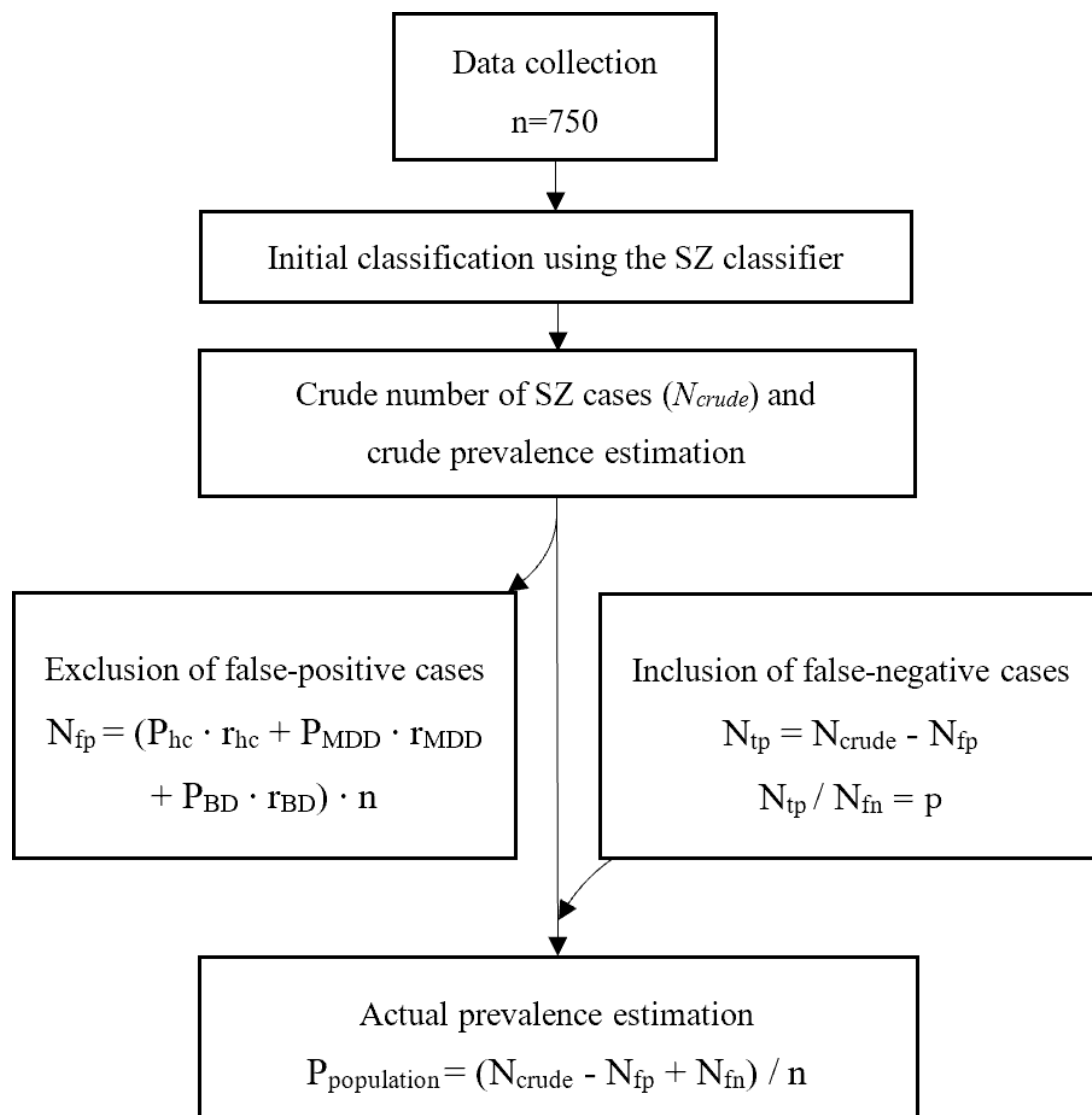
where N_{tp} represents the cases predicted by the SZ classifier, N_{crude} represents the crude number of schizophrenia cases estimated by the SZ classifier, N_{fn} represents the false-negative cases, and p represents the proportion of true-positive cases to false-negative cases in the external validation, which was 46:15 [15].

Finally, the prevalence of schizophrenia in the population could be calculated by

$$P_{population} = (N_{crude} - N_{fp} + N_{fn})/n$$

where $P_{population}$ represents the prevalence of schizophrenia in the population (Figure 1).

Figure 1. Flowchart illustrating the data analysis and refinement process for estimating the prevalence of schizophrenia in the general Japanese population. Data were collected from 750 participants during November 20 to 26, 2023. Participants' demographic information, health-related background, and physical, psychiatric, and social comorbidities were assessed using the schizophrenia classification model to initially estimate the crude number of schizophrenia cases and crude prevalence. The refinement process involved excluding false-positive cases and including false-negative cases, resulting in an actual estimation of schizophrenia prevalence in the general population of Japan.



Statistical Analysis

The 95% CIs were calculated from 10,000 bootstrap replicates. Statistical analyses were performed using Python (version 3.8; Python Software Foundation), with a Jupyter Notebook (Project Jupyter) as a coding notebook.

Ethical Considerations

The study protocol was reviewed and approved by the institutional review board at Fujita Health University (HM23-253). All procedures were performed following the ethical standards of the Declaration of Helsinki. The internet survey agency respected the Act on the Protection of Personal Information in Japan. Informed consent was obtained at the initial web page of the survey, where the purpose of the survey and the instructions to the respondents were displayed. Only those who had read the survey's purpose and instructions and provided informed consent by clicking the agreement button could proceed with the survey. Participants

for the external validation of the SZ classifier were diagnosed by experienced psychiatrists. The data were anonymous and deidentified before being provided to researchers. This was a voluntary survey. Participants received Rakuten Points, which can be used for Rakuten services such as web-based or offline shopping. No image data were collected so that there were no concerns about identifying individual participants.

Results

Of the 750 samples analyzed, 62 were classified as schizophrenia cases (N_{crude}). The crude prevalence of schizophrenia in the general population of Japan, as estimated using the SZ classifier, was 8.3% (95% CI 6.6%-10.1%; Table 1). Among these 62 cases, 53 were presumed to be false positives, and 3 were presumed to be false negatives. After adjustments, the actual prevalence of schizophrenia in the general population was estimated to be 1.6% (95% CI 0.7%-2.5%; Table 2).

Table 1. Results of crude estimates of schizophrenia cases by the schizophrenia classification model (SZ classifier), crude schizophrenia prevalence, false-positive estimates, and true-positive estimates.

	Cases, n	Prevalence, %
Crude estimates of predicted schizophrenia by the SZ classifier	62	8.3 (6.6-10.1) ^a
False positive: no mental disorders misclassified as schizophrenia	18	2.5
False positive: patients with MDD ^b misclassified as schizophrenia	33	4.4
False positive: patients with BD ^c misclassified as schizophrenia	2	0.2
True positive: patients with schizophrenia who were accurately classified as schizophrenia cases	9	1.2

^a95% CI value.
^bMDD: major depressive disorder.
^cBD: bipolar disorder.

Table 2. Confusion matrix and adjusted prevalence of schizophrenia in the general population of Japan.

	Adjusted cases estimation	
	Schizophrenia	NonSchizophrenia
Predicted cases, n		
Schizophrenia	9	53
Nonschizophrenia	3	685
Total	12	738
Prevalence, % (95%CI)	1.6 (0.7-2.5)	— ^a

^aNot applicable.

Discussion

Principal Results

Our findings indicate that the estimated prevalence of schizophrenia in the general population of Japan using the SZ classifier was 1.6% (95% CI 0.7%-2.5%). This estimate is slightly higher than those in previously published figures [5,6]. This discrepancy could be attributed to several reasons, including the possibility that our SZ classifier estimates lifetime prevalence rather than point prevalence, a potentially more comprehensive disease classification methodology, and calculating the actual prevalence by referencing various prevalence data from diverse published sources.

To ensure the validity of the data used in our study, 3 main points were considered. The first point is the trustworthiness of the data source. The data were collected and managed by Rakuten Insight Panel, a large-scale web-based survey panel. The panel's quality has been maintained by company procedures and has been widely used by research institutions in Japan for population-based studies [16]. The second point is the appropriateness of the data used for model development. All feature variables were decided both through a literature review [17] and studies demonstrating that all these variables effectively differentiate between individuals with and those without schizophrenia [14,15,18]. Additionally, all the schizophrenia cases were defined by using a screening protocol instead of relying solely on the self-reported illness conditions. Schizophrenia cases had to meet the following four criteria: (1) currently having schizophrenia only, schizophrenia and migraine, schizophrenia and a sleep disorder, or schizophrenia, migraine, and a sleep disorder; (2) having experienced auditory hallucinations lasting more than 1 month; (3) not having used stimulants or other illegal

drugs and not having been an alcoholic; and (4) having experienced the first auditory hallucination lasting more than 1 month at less than 60 years of age [18]. Third, the high performance of the developed model relies on high-quality data. The robustness and generalizability of the model were confirmed through good internal validation (AUC=0.86) [14] and external validation results (sensitivity=0.75) [15].

The SZ classifier predicts schizophrenia status using various feature variables. Each feature has been reported as being significantly associated with schizophrenia in previous studies [14,15,17,18]. By training the model with these features and applying data from the study participants, the probability of schizophrenia can be calculated. Although this probability may not be highly interpretable for diagnosis on an individual level, it reliably represents the proportion of specific cases within a target group. When the model demonstrates good generalizability, this proportion can be extended to estimate prevalence by considering the false positives and false negatives on a population level.

According to the feature importance ranking estimated by the model, social comorbidities, such as household income, type of employment, educational background, and type of occupation, play significant roles [14]. These factors may have been influenced by previous psychiatric conditions, which could have had long-term effects on the data used in this analysis. For instance, individuals diagnosed with schizophrenia may have had to stop attending school or leave their jobs, affecting their current income and occupation [22]. Consequently, the model might have predicted them as having schizophrenia even though they had recovered, which is consistent with the lifetime prevalence estimates. Several studies have shown that the prevalence of schizophrenia increases when it extends from one point to the lifetime of an individual [4,23].

To the best of our knowledge, this is the first prevalence study to apply an ANN model as a schizophrenia case classifier using web-based survey data from the general population of Japan. ANNs, which are computational models inspired by the inner processes of the human brain, are increasingly used in public health research [24,25]. Technically, ANNs are excellent at capturing complex and nonlinear relationships within large datasets. Hence, models constructed using ANNs perform better than classical methods based on linear equations [26,27]. Several studies have proposed ANN models or machine learning techniques to detect and classify schizophrenia [28,29]. These studies developed models using large amounts of clinical data (ie, magnetic resonance images, positron emission tomography scans, electroencephalography, and subjective patient posture interpretation). These studies have applied machine learning techniques to improve detection and diagnosis and aid in clinical decision-making. Bae et al [30] demonstrated the use of machine learning to detect signs of schizophrenia by analyzing social media texts, supporting the role of machine learning models in improving symptoms monitoring and potentially early detection of schizophrenia.

The use of ANN models to estimate prevalence has been reported in previous studies. For example, suicidal ideation risk prediction models were developed based on demographic, behavioral, and psychological data obtained from computer-based questionnaires. These models were then used to estimate the prevalence of suicidal ideation among Chinese college students. The models demonstrated a prediction accuracy of over 91.7%, and the AUC score increased to 0.96 [31]. Compared with previous studies, our SZ classifier has undergone external validation to demonstrate its generalizability [15]. This study supports the use of machine learning-based methods for prevalence estimation.

This study design may provide evidence to overcome biases encountered in traditional self-identifying surveys due to stigma, poor insight, and lack of access to health professionals experienced in conventional self-identifying surveys [32,33]. Stocking et al [34] reported that they first used a mathematical model to explore the correlation between several comorbidities and a specific illness in a small sample, and then applied this correlation to a larger population to identify previously underreported cases. They noted that this approach is effective for addressing the underreporting of various chronic conditions associated with social stigma, challenges in detection and treatment, or inconsistent treatment engagement (eg, mental disorders, obesity, hypertension, sleep disorders, etc). In this study, schizophrenia cases were estimated by incorporating general information, including demographics, health-related backgrounds, and physical, psychological, and social comorbidities, to calculate the likelihood of schizophrenia. This approach encompasses a broader range of patients in the general population than previous methodologies that relied on clinical records or typical schizophrenia symptoms. In contrast to a diagnosis that depends on the self-reporting of typical schizophrenia symptoms, which requires participants to openly acknowledge symptoms that may suggest

schizophrenia, our methodology provides a more objective assessment. Patients with schizophrenia are often stigmatized and not widely accepted in society, leading to embarrassment and hesitation in disclosing their symptoms and avoiding negative perceptions. Stigmatizing attitudes related to mental illness have been reported to prevent individuals from seeking counseling [35] and harm the self-esteem of people with serious mental illnesses [36]. Additionally, the wording of questionnaires that focus on negative aspects may cause respondents to feel negative and skip or conceal their actual symptoms, resulting in underreported results in self-identifying surveys [37,38]. By using general information, our SZ classifier mitigated these biases, potentially providing a more comprehensive estimate of patients with schizophrenia in Japan.

The SZ classifier has the advantage of potentially detecting patients with schizophrenia who do not perceive subjective symptoms. Hoffman et al [39] reported that a common reason why people with schizophrenia lack insight into their illness is that their ability to objectively assess their own pathological experiences is impaired. By using objective indicators such as demographics, health-related backgrounds, and physical, psychological, and social comorbidities, the SZ classifier showed a potential ability of detecting patients who do not perceive subjective symptoms and estimating the prevalence of schizophrenia in the population. This study may offer unique advantages over previous estimation methods based solely on medical data.

The higher prevalence estimated in our study suggests that schizophrenia might be more common than previously believed in the Japanese population. This finding confirmed our hypothesis regarding the underestimation of schizophrenia cases in the Japanese population. These findings have significant implications for public health management and resource allocation. Greater knowledge about the actual prevalence of schizophrenia can result in enhanced mental health services and increased support systems for managing this public health concern.

Limitations

Our study has several limitations. First, the actual prevalence of schizophrenia was calculated by referencing various prevalence data from diverse published sources, including the prevalence of individuals who did not have any mental disorder, the prevalence of those who had MDD, and the prevalence of those who had BD. Using data from different studies, which may have used different methodologies and population samples, causes uncertainty in the data, thereby affecting the estimated values. Second, although the data used in this study were collected from an extensive web-based survey that efficiently integrates large amounts of data, it may introduce a selection bias toward those with internet access. Those who are less likely to participate in web-based surveys, such as individuals without internet access, institutionalized patients, patients with severe comorbidities, older adults, and poor people, might be underrepresented. Third, the SZ classifier, designed to capture a broad range of patients, may include individuals with subclinical symptoms or comorbid

conditions similar to schizophrenia. Despite efforts to rule out incorrectly classified cases, there is a chance that other mental disorders will not be sufficiently considered. This inclusion may lead to an overestimation of the true prevalence of schizophrenia. Fourth, for mental disorders comorbid with schizophrenia symptoms (eg, schizoaffective disorder, characterized by symptoms of both schizophrenia and mood disorders, etc), these cases were not included in the SZ classifier validation. Also, due to the complexity of mental disorders, the diagnosis of a disease without confirmation from a physician may seem arbitrary. Therefore, these results should be interpreted with caution.

Conclusions

This study applied a pretrained SZ classifier to estimate the prevalence of schizophrenia in a Japanese population. The

estimated results are slightly higher than those of previous studies. Considering the shortcomings of focusing on a single type of mental disorder and the need to increase the sample size, this study shows promise for prevalence-estimation research. Potential future research directions may focus on developing models to classify other mental disorders, such as BD and MDD, as well as estimate prevalence. These findings are expected to support the versatility of pretrained models in epidemiological research.

Acknowledgments

This study was funded by the Ministry of Health, Labour and Welfare, Japan (JPMH21GC1018). The findings and conclusions of this article are the sole responsibility of the authors and do not represent the official views of the research funders. The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this manuscript.

Data Availability

The datasets analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: PC, YH, and AO

Data curation: MM and AO

Data interpretation: PC and YH

Data analysis: PC and YH

Investigation: MM and AO

Methodology: AO and YH

Supervision: AO

Writing—original draft: PC

Writing—review and editing: YH, KS, TK, YL, ST, NI, and AO

Conflicts of Interest

None declared.

Multimedia Appendix 1

Inclusion criteria, variable definition, and model validation results.

[\[DOCX File \(Microsoft Word File\), 36 KB-Multimedia Appendix 1\]](#)

References

1. Marder SR, Cannon TD. Schizophrenia. *N Engl J Med*. Oct 31, 2019;381(18):1753-1761. [doi: [10.1056/NEJMra1808803](https://doi.org/10.1056/NEJMra1808803)] [Medline: [31665579](https://pubmed.ncbi.nlm.nih.gov/31665579/)]
2. World Health Organization. Schizophrenia. Jan 10, 2022. URL: <https://www.who.int/news-room/fact-sheets/detail/schizophrenia> [Accessed 2024-05-17]
3. 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*. Feb 2022;9(2):137-150. [doi: [10.1016/S2215-0366\(21\)00395-3](https://doi.org/10.1016/S2215-0366(21)00395-3)] [Medline: [35026139](https://pubmed.ncbi.nlm.nih.gov/35026139/)]
4. Simeone JC, Ward AJ, Rotella P, Collins J, Windisch R. An evaluation of variation in published estimates of schizophrenia prevalence from 1990—2013: a systematic literature review. *BMC Psychiatry*. Aug 12, 2015;15:193. [doi: [10.1186/s12888-015-0578-7](https://doi.org/10.1186/s12888-015-0578-7)] [Medline: [26263900](https://pubmed.ncbi.nlm.nih.gov/26263900/)]
5. Sado M, Inagaki A, Koreki A, et al. The cost of schizophrenia in Japan. *Neuropsychiatr Dis Treat*. 2013;9:787-798. [doi: [10.2147/NDT.S41632](https://doi.org/10.2147/NDT.S41632)] [Medline: [23785238](https://pubmed.ncbi.nlm.nih.gov/23785238/)]

6. Baba K, Guo W, Chen Y, Nosaka T, Kato T. Burden of schizophrenia among Japanese patients: a cross-sectional National Health and Wellness Survey. *BMC Psychiatry*. Jun 18, 2022;22(1):410. [doi: [10.1186/s12888-022-04044-5](https://doi.org/10.1186/s12888-022-04044-5)] [Medline: [35717149](https://pubmed.ncbi.nlm.nih.gov/35717149/)]
7. Rodríguez-Rivas ME, Cangas AJ, Cariola LA, Varela JJ, Valdebenito S. Innovative technology-based interventions to reduce stigma toward people with mental illness: systematic review and meta-analysis. *JMIR Serious Games*. May 30, 2022;10(2):e35099. [doi: [10.2196/35099](https://doi.org/10.2196/35099)] [Medline: [35635744](https://pubmed.ncbi.nlm.nih.gov/35635744/)]
8. Gray AJ. Stigma in psychiatry. *J R Soc Med*. Feb 2002;95(2):72-76. [doi: [10.1177/014107680209500205](https://doi.org/10.1177/014107680209500205)] [Medline: [11823548](https://pubmed.ncbi.nlm.nih.gov/11823548/)]
9. Kamaradova D, Latalova K, Prasko J, et al. Connection between self-stigma, adherence to treatment, and discontinuation of medication. *Patient Prefer Adherence*. 2016;10:1289-1298. [doi: [10.2147/PPA.S99136](https://doi.org/10.2147/PPA.S99136)] [Medline: [27524884](https://pubmed.ncbi.nlm.nih.gov/27524884/)]
10. Vrbova K, Prasko J, Holubova M, et al. Self-stigma and schizophrenia: a cross-sectional study. *Neuropsychiatr Dis Treat*. 2016;12:3011-3020. [doi: [10.2147/NDT.S120298](https://doi.org/10.2147/NDT.S120298)] [Medline: [27920538](https://pubmed.ncbi.nlm.nih.gov/27920538/)]
11. Konstantakopoulos G. Insight across mental disorders: a multifaceted metacognitive phenomenon. *Psychiatriki*. 2019;30(1):13-16. [doi: [10.22365/jpsych.2019.301.13](https://doi.org/10.22365/jpsych.2019.301.13)] [Medline: [31115349](https://pubmed.ncbi.nlm.nih.gov/31115349/)]
12. Ando S, Yamaguchi S, Aoki Y, Thornicroft G. Review of mental-health-related stigma in Japan. *Psychiatry Clin Neurosci*. Nov 2013;67(7):471-482. [doi: [10.1111/pcn.12086](https://doi.org/10.1111/pcn.12086)] [Medline: [24118217](https://pubmed.ncbi.nlm.nih.gov/24118217/)]
13. Goldner EM, Hsu L, Waraich P, Somers JM. Prevalence and incidence studies of schizophrenic disorders: a systematic review of the literature. *Can J Psychiatry*. Nov 2002;47(9):833-843. [doi: [10.1177/070674370204700904](https://doi.org/10.1177/070674370204700904)] [Medline: [12500753](https://pubmed.ncbi.nlm.nih.gov/12500753/)]
14. He Y, Matsunaga M, Li Y, et al. Classifying schizophrenia cases by artificial neural network using Japanese web-based survey data: case-control study. *JMIR Form Res*. Nov 15, 2023;7:e50193. [doi: [10.2196/50193](https://doi.org/10.2196/50193)] [Medline: [37966882](https://pubmed.ncbi.nlm.nih.gov/37966882/)]
15. He Y, Sakuma K, Kishi T, et al. External validation of a machine learning model for schizophrenia classification. *J Clin Med*. May 17, 2024;13(10):2970. [doi: [10.3390/jcm13102970](https://doi.org/10.3390/jcm13102970)] [Medline: [38792511](https://pubmed.ncbi.nlm.nih.gov/38792511/)]
16. Rakuten Insight, Inc. Rakuten Insight. URL: <https://insight.rakuten.co.jp/en/aboutus.html> [Accessed 2024-05-30]
17. He Y, Tanaka A, Kishi T, et al. Recent findings on subjective well-being and physical, psychiatric, and social comorbidities in individuals with schizophrenia: a literature review. *Neuropsychopharmacol Rep*. Dec 2022;42(4):430-436. [doi: [10.1002/npr2.12286](https://doi.org/10.1002/npr2.12286)] [Medline: [35916310](https://pubmed.ncbi.nlm.nih.gov/35916310/)]
18. Matsunaga M, Li Y, He Y, et al. Physical, psychiatric, and social comorbidities of individuals with schizophrenia living in the community in Japan. *Int J Environ Res Public Health*. Feb 28, 2023;20(5):4336. [doi: [10.3390/ijerph20054336](https://doi.org/10.3390/ijerph20054336)] [Medline: [36901345](https://pubmed.ncbi.nlm.nih.gov/36901345/)]
19. Nishi D, Ishikawa H, Kawakami N. Prevalence of mental disorders and mental health service use in Japan. *Psychiatry Clin Neurosci*. Aug 2019;73(8):458-465. [doi: [10.1111/pcn.12894](https://doi.org/10.1111/pcn.12894)] [Medline: [31141260](https://pubmed.ncbi.nlm.nih.gov/31141260/)]
20. Hoshino E, Ohde S, Rahman M, Takahashi O, Fukui T, Deshpande GA. Variation in somatic symptoms by Patient Health Questionnaire-9 depression scores in a representative Japanese sample. *BMC Public Health*. Dec 27, 2018;18(1):1406. [doi: [10.1186/s12889-018-6327-3](https://doi.org/10.1186/s12889-018-6327-3)] [Medline: [30587173](https://pubmed.ncbi.nlm.nih.gov/30587173/)]
21. Kato T. Bipolar disorder: from pathophysiology to treatment. *Juntendo Iji Zasshi*. 2022;68(1):17-24. [doi: [10.14789/jmj.JMJ21-0026-R](https://doi.org/10.14789/jmj.JMJ21-0026-R)] [Medline: [38911011](https://pubmed.ncbi.nlm.nih.gov/38911011/)]
22. Hakulinen C, McGrath JJ, Timmerman A, et al. The association between early-onset schizophrenia with employment, income, education, and cohabitation status: nationwide study with 35 years of follow-up. *Soc Psychiatry Psychiatr Epidemiol*. Nov 2019;54(11):1343-1351. [doi: [10.1007/s00127-019-01756-0](https://doi.org/10.1007/s00127-019-01756-0)] [Medline: [31456027](https://pubmed.ncbi.nlm.nih.gov/31456027/)]
23. Bhugra D. The global prevalence of schizophrenia. *PLoS Med*. May 2005;2(5):e151. [doi: [10.1371/journal.pmed.0020151](https://doi.org/10.1371/journal.pmed.0020151)] [Medline: [15916460](https://pubmed.ncbi.nlm.nih.gov/15916460/)]
24. Lanillos P, Oliva D, Philippsen A, Yamashita Y, Nagai Y, Cheng G. A review on neural network models of schizophrenia and autism spectrum disorder. *Neural Netw*. Feb 2020;122:338-363. [doi: [10.1016/j.neunet.2019.10.014](https://doi.org/10.1016/j.neunet.2019.10.014)] [Medline: [31760370](https://pubmed.ncbi.nlm.nih.gov/31760370/)]
25. Razavi M, Ziyadidegan S, Mahmoudzadeh A, et al. Machine learning, deep learning, and data preprocessing techniques for detecting, predicting, and monitoring stress and stress-related mental disorders: scoping review. *JMIR Ment Health*. Aug 21, 2024;11:e53714. [doi: [10.2196/53714](https://doi.org/10.2196/53714)] [Medline: [39167782](https://pubmed.ncbi.nlm.nih.gov/39167782/)]
26. Masegosa AR, Cabañas R, Langseth H, Nielsen TD, Salmerón A. Probabilistic models with deep neural networks. *Entropy (Basel)*. Jan 18, 2021;23(1):117. [doi: [10.3390/e23010117](https://doi.org/10.3390/e23010117)] [Medline: [33477544](https://pubmed.ncbi.nlm.nih.gov/33477544/)]
27. He Y, Sun Q, Matsunaga M, Ota A. Can feature structure improve model's precision? A novel prediction method using artificial image and image identification. *JAMIA Open*. Apr 2024;7(1):ooae012. [doi: [10.1093/jamiaopen/ooae012](https://doi.org/10.1093/jamiaopen/ooae012)] [Medline: [38348347](https://pubmed.ncbi.nlm.nih.gov/38348347/)]

28. Lai JW, Ang CKE, Acharya UR, Cheong KH. Schizophrenia: a survey of artificial intelligence techniques applied to detection and classification. *Int J Environ Res Public Health*. Jun 5, 2021;18(11):6099. [doi: [10.3390/ijerph18116099](https://doi.org/10.3390/ijerph18116099)] [Medline: [34198829](https://pubmed.ncbi.nlm.nih.gov/34198829/)]
29. Verma S, Goel T, Tanveer M, Ding W, Sharma R, Murugan R. Machine learning techniques for the schizophrenia diagnosis: a comprehensive review and future research directions. *J Ambient Intell Human Comput*. May 2023;14(5):4795-4807. [doi: [10.1007/s12652-023-04536-6](https://doi.org/10.1007/s12652-023-04536-6)]
30. Bae YJ, Shim M, Lee WH. Schizophrenia detection using machine learning approach from social media content. *Sensors (Basel)*. Sep 3, 2021;21(17):5924. [doi: [10.3390/s21175924](https://doi.org/10.3390/s21175924)] [Medline: [34502815](https://pubmed.ncbi.nlm.nih.gov/34502815/)]
31. Liao S, Wang Y, Zhou X, et al. Prediction of suicidal ideation among Chinese college students based on radial basis function neural network. *Front Public Health*. 2022;10:1042218. [doi: [10.3389/fpubh.2022.1042218](https://doi.org/10.3389/fpubh.2022.1042218)] [Medline: [36530695](https://pubmed.ncbi.nlm.nih.gov/36530695/)]
32. Supina AL, Patten SB. Self-reported diagnoses of schizophrenia and psychotic disorders may be valuable for monitoring and surveillance. *Can J Psychiatry*. Mar 2006;51(4):256-259. [doi: [10.1177/070674370605100407](https://doi.org/10.1177/070674370605100407)] [Medline: [16629350](https://pubmed.ncbi.nlm.nih.gov/16629350/)]
33. Gould F, McGuire LS, Durand D, et al. Self-assessment in schizophrenia: accuracy of evaluation of cognition and everyday functioning. *Neuropsychology*. Sep 2015;29(5):675-682. [doi: [10.1037/neu0000175](https://doi.org/10.1037/neu0000175)] [Medline: [25643212](https://pubmed.ncbi.nlm.nih.gov/25643212/)]
34. Stocking A, Duncan I, Huynh N. Estimating underdiagnosis of patients in chronically ill populations. *N Am Actuar J*. Dec 18, 2023;1-10. [doi: [10.1080/10920277.2023.2281471](https://doi.org/10.1080/10920277.2023.2281471)]
35. Vogel DL, Wade NG, Hackler AH. Perceived public stigma and the willingness to seek counseling: the mediating roles of self-stigma and attitudes toward counseling. *J Couns Psychol*. 2007;54(1):40-50. [doi: [10.1037/0022-0167.54.1.40](https://doi.org/10.1037/0022-0167.54.1.40)]
36. Link BG, Struening EL, Neese-Todd S, Asmussen S, Phelan JC. Stigma as a barrier to recovery: the consequences of stigma for the self-esteem of people with mental illnesses. *Psychiatr Serv*. Dec 2001;52(12):1621-1626. [doi: [10.1176/appi.ps.52.12.1621](https://doi.org/10.1176/appi.ps.52.12.1621)] [Medline: [11726753](https://pubmed.ncbi.nlm.nih.gov/11726753/)]
37. Subramaniam M, Abdin E, Vaingankar JA, et al. Lifetime prevalence and correlates of schizophrenia and other psychotic disorders in Singapore. *Front Psychiatry*. 2021;12:650674. [doi: [10.3389/fpsy.2021.650674](https://doi.org/10.3389/fpsy.2021.650674)] [Medline: [33776823](https://pubmed.ncbi.nlm.nih.gov/33776823/)]
38. Karam EG, Sampson N, Itani L, et al. Under-reporting bipolar disorder in large-scale epidemiologic studies. *J Affect Disord*. Apr 2014;159:147-154. [doi: [10.1016/j.jad.2014.01.011](https://doi.org/10.1016/j.jad.2014.01.011)] [Medline: [24679403](https://pubmed.ncbi.nlm.nih.gov/24679403/)]
39. Hoffman RE, McGlashan TH. Corticocortical connectivity, autonomous networks, and schizophrenia. *Schizophr Bull*. 1994;20(2):257-261. [doi: [10.1093/schbul/20.2.257](https://doi.org/10.1093/schbul/20.2.257)] [Medline: [8085129](https://pubmed.ncbi.nlm.nih.gov/8085129/)]

Abbreviations

- ANN:** artificial neural network
AUC: area under the receiver operating characteristic curve
BD: bipolar disorder
MDD: major depressive disorder
OCD: obsessive-compulsive disorder
SZ classifier: schizophrenia classification model

Edited by Amaryllis Mavragani; peer-reviewed by Doaa Abdel-Hady, Sokaina El Khamlichi; submitted 10.09.2024; final revised version received 05.01.2025; accepted 07.01.2025; published 29.01.2025

Please cite as:

Choomung P, He Y, Matsunaga M, Sakuma K, Kishi T, Li Y, Tanihara S, Iwata N, Ota A
Estimating the Prevalence of Schizophrenia in the General Population of Japan Using an Artificial Neural Network-Based Schizophrenia Classifier: Web-Based Cross-Sectional Survey
JMIR Form Res 2025;9:e66330
URL: <https://formative.jmir.org/2025/1/e66330>
doi: [10.2196/66330](https://doi.org/10.2196/66330)

© Pichsinee Choomung, Yupeng He, Masaaki Matsunaga, Kenji Sakuma, Taro Kishi, Yuanying Li, Shinichi Tanihara, Nakao Iwata, Atsuhiko Ota. Originally published in *JMIR Formative Research* (<https://formative.jmir.org>), 29.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Formative Research*, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.