VENN: Journal of Sustainable Innovation on Education, Mathematics and Natural Sciences
ISSN: 2964-867X
Journal Homepage: https://pustaka.my.id/journals/venn/index

Estimating Odds Ratios of Factors Associated with Stimulant Use among Outpatients: A Bayesian Logistic Regression Case Study at BNNK Palopo, Indonesia

Besse Helmi Mustawinar[1], Yuliani[2], Dwi Risky Arifanti[3], Ikes Dwiastuti[4], Tsabitah Fausiyah Azzahra[5], Nasti Novyandini[6]

Department of Mathematics, Faculty of Sciences, Universitas Cokroaminoto Palopo, Indonesia[1,2,5,6]
Department of Mathematics Education Study, Tarbiyah and Teacher Training Faculty, UIN Palopo, Indonesia[3]
Badan Narkotika Nasional (BNN) Kota Palopo, Indonesia[4]
emmy.emm92@gmail.com[1], *yulimath27@gmail.com[2]*, *dwiriskyarifanti@uinpalopo.ac.id[3]*
ikesdwiastuti@gmail.com[4], *tarahtahzahra09@gmail.com[5]*, *nastinovyandini@gmail.com[6]*

Corresponding Author: Besse Helmi Mustawinar, E-mail: emmy.emm92@gmail.com

Article Information
Received : 28 August 2026
Reviced : 22 November 2026
Accepted : 26 November 2026
Published : 26 November 2026

ABSTRACT
Identifying factors associated with the type of substance used by patients can help the Badan Narkotika Nasional Kota (BNNK) Palopo plan targeted rehabilitation services. However, small clinical datasets with an unbalanced outcome often yield unstable estimates under maximum likelihood logistic regression. This study estimated the odds ratios (ORs) of four patient characteristics (current age, occupational status, age at first use, and environmental influence) for the use of stimulants versus non-stimulants among outpatients at BNNK Palopo. We conducted a retrospective analysis of secondary records of 74 outpatients with complete assessment data (55 coded as stimulant use and 19 coded as non-stimulant). Because the smaller outcome category contained only 19 patients, Bayesian logistic regression with weakly informative priors centred as zero was used to regularise the estimates, fitted using Hamiltonian Monte Carlo (four chains, 4,000 iterations each, 1,000 warmup). Convergence was satisfactory ($\hat{R}$ = 1.00 for all parameters). Age at first use was the only predictor whose 95% credible interval excluded 1 (posterior median OR = 1.26 per one-year increase; 95% CrI 1.08–1.51), indicating that later initiation was associated with higher odds of stimulant use in this sample. The credible intervals for current age, occupational status, and environmental influence included 1. Given the small sample, the single-site design, and the dependence of Bayesian estimates on prior choice, these findings are exploratory and should not be generalized beyond this outpatient sample.

KEYWORDS
Bayesian Logistic Regression; Initiation Age; Odds Ratio; Stimulants

ARTICLE DOI:
https://doi.org/10.53696/venn.v5i6.663

How to Cite: Mustawinar, B. H., Yuliani, Y., Arifanti, D. R., Dwiastuti, I., Azzahra, T. F., & Novyandini, N. (2026). Estimating Odds Ratios of Factors Associated with Stimulant Use among Outpatients: A Bayesian Logistic Regression Case Study at BNNK Palopo, Indonesia. Venn: Journal of Sustainable Innovation on Education, Mathematics and Natural Sciences, 5(6), 1403–1413. https://doi.org/10.53696/venn.v5i6.663

Copyright: © 2026 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/).

A. Introduction

Drug misuse remains a major public health concern in Indonesia. The national survey by the Badan Narkotika Nasional (BNN, 2025) reported that prevalence rose to 2.11% (4.15 million people) from 1.73% in 2023, with the productive-age population (15–64 years) most affected. National figures, however, do not describe the pattern of substance use among patients in individual districts. In 2025, BNNK Palopo whose working area covers Palopo, Luwu, North Luwu, and East Luwu, assessed 250 suspected drug misusers, of whom 233 underwent rehabilitation. According to data from the Integrated Assessment Team and the Jaya Sakti Primary Clinic, these individuals were directed into outpatient, inpatient, or post-rehabilitation programs, all of which are provided free of charge and without any accompanying criminal legal proceedings (Nandini, 2025). These administrative data show that clients are being treated in this region, but they do not describe the distribution of substances among treated patients, and population-level epidemiological data for this area are limited. Understanding which patient characteristics are associated with the use of stimulant substances may help BNNK Palopo tailor its rehabilitation services.

The analysis of demographic and psychosocial predictors, especially age and employment status, is crucial for understanding the landscape of individual risk in the context of narcotics rehabilitation (Rioux et al., 2018; Sinha et al., 2003). Age, as a central variable, can describe a person's physical development and brain functions. . The clinical literature has identified the age of First Drug Use as an important correlate of later dependence and problematic use (Yamani & Alkhatib, 2025). Early exposure to addictive substances, particularly during brain development, can result in lasting changes in brain structure and trigger long-term behavioral changes (Nath et al., 2022). These effects can increase the risk of future dependence. This literature, however, mostly concerns the risk or severity of dependence rather than the type of substance used. Evidence relating initiation age to the pattern of substances used is more limited and depends on the comparison group and setting.

Empirical evidence, consistently observed through national surveys such as the National Survey on Drug Use and Health (NSDUH), which collects data on the Age of First Use across several substance categories, including alcohol, marijuana, cocaine, and prescription drugs, shows that the majority of individuals begin drug use before reaching full adulthood (Knopf, 2023b, 2023a; Rosner et al., 2020; Sathe et al., 2022; *The National Survey on Drug Use and Health (NSDUH)*, 2003). Studies of treatment-seeking populations have mostly described specific groups rather than comparing substance types: for example, U.S. treatment admissions (TEDS-A) of older adults with heroin or psychostimulant use (Choi et al., 2021) and a scale-development study of adolescent amphetamine users in Thailand (Uthis et al., 2025). Evidence comparing initiation age between types of substance is therefore limited. Whether initiation age differs between outpatients who use stimulants and those who do not remains unclear in Indonesian treatment settings.

Although initiation age is an important correlate of dependence risk, its influence often interacts with socioeconomic stability, as reflected in employment status. According to Marie Johada's latent deprivation model, employment provides routine, social identity, and support networks, serving as protective factors against addictive behavior (Paul et al., 2023). Conversely, the unemployed group showed a significantly higher rate of Substance Use Disorder (SUD) The COSINUS cohort study showed that participants engaged in productive activities were less likely to report problematic stimulant use (adjusted odds ratio 0.60). Job instability can exacerbate vulnerability among individuals with a history of early initiation, creating a double-risk burden that complicates their recovery (Roux et al., 2024). These studies address the risk of dependence or problematic use rather than the type of substance used, so their relevance to substance use patterns is indirect.

Apart from occupational aspects, peer influence is an equally crucial social determinant. Friendship circles often serve as the main entry point and a factor that perpetuates substance use through social contagion mechanisms. In the context of stimulants such as methamphetamine, use often occurs in groups as part of social activities or to boost self-confidence during social interactions (Nath et al., 2022; Rahman et al., 2016). The normalisation of substance use within peer circles decreases individuals' perception of risk, making peer pressure a strong predictor of continued substance use among adolescents (Pomrenze et al., 2022). In contrast, the family environment functions as the primary unit of socialization, which can serve as either a protective or risk factor. Family disharmony, lack of parental supervision, and the presence of other family members with a history of addiction can create an environment permissive to drug use (Shoaa Kazemi et al., 2025).

Methodologically, risk factor analysis of clinical data often faces limitations, such as small sample sizes and unbalanced data distributions. Many analyses of clinical data rely on frequentist approaches, such as Maximum Likelihood Estimation (MLE) in logistic regression. However, MLE has limitations when dealing with small samples, where parameter estimates can be unstable and standard errors imprecise. This can lead to less accurate conclusions in public health policymaking (Pavlou et al., 2024; Zhu et al., 2024). To address these limitations, this study applied Bayesian logistic regression with weakly informative priors, which can regularise the estimates and express uncertainty directly through posterior distributions, although the results depend on the chosen priors and should be examined in sensitivity analyses (Pavlou et al., 2024; J. Zhu et al., 2024). The Hamiltonian Monte Carlo (HMC) algorithm allows the posterior distribution of the odds ratios to be sampled efficiently (Sun & Reddy, 2013; Zhao et al., 2024).

Previous studies have examined age at first use, employment, and the social environment mainly in relation to the risk of dependence or problematic substance use. Evidence on how these demographic and environmental factors are associated with the type of substance used is scarce, and, to our knowledge, no estimates of these associations are available for the outpatient population of BNNK Palopo. This study addresses that gap by estimating the odds ratios of current age, occupational status, age at first use, and environmental influence for stimulant use versus non-stimulants use patterns among outpatients at BNNK Palopo, using Bayesian logistic regression. The findings are expected to provide an initial empirical basis for planning rehabilitation services at BNNK Palopo and for designing larger studies in the region.

B. Methods

This study employed a quantitative, retrospective case study design to estimate the factors associated with stimulant use among outpatients undergoing drug rehabilitation. This research was conducted at the BNNK Palopo in South Sulawesi, Indonesia. Data collection and extraction were performed for one year. Because this study involved human clinical data, ethical clearance and administrative permission to access secondary medical records were obtained directly from the Internal Review Board of BNNK Palopo. All patient identities were anonymized prior to data extraction to ensure strict confidentiality.

This study focused on outpatients attending rehabilitation programs at BNNK Palopo. Researchers purposively included clients with complete assessment records; therefore, no data were missing for the variables analysed (complete-case analysis). The final sample comprised 74 outpatients: 55 were coded as stimulant users (methamphetamine as the only recorded substance) and 19 as non-stimulant (all other recorded patterns of use: non-methamphetamine substances only, n = 6, or combinations that included methamphetamine, n = 13). These forms have been proven to work well for checking addiction severity. The details of the data are presented in Table 1.

Table 1. Research Variables and Operational Definitions

Variable	Variabel Name	Categorization	Scale
Y	Substance Preference	0: Non-stimulants; 1: Stimulants	Dichotomous
X1	Current Age	The biological age of the respondent at the time of the research (in years)	Ratio
X2	Occupational Status	0: Unemployed; 1: Student; 2: Self-employed; 3: Civil Servant	Nominal
X3	Initiation Age	The age at which the respondent first consumed illicit substances (in years)	Ratio
X4	Environment Influence	0: Family; 1: Peers/Friends; 2: Community/Workplace	Nominal

The procedure of this study is illustrated in Figure 1. The flow begins with the preprocessing of clinical data, then specifies the Bayesian model using weakly informative priors, and proceeds to the simulation using the HMC algorithm to produce a stable posterior distribution.

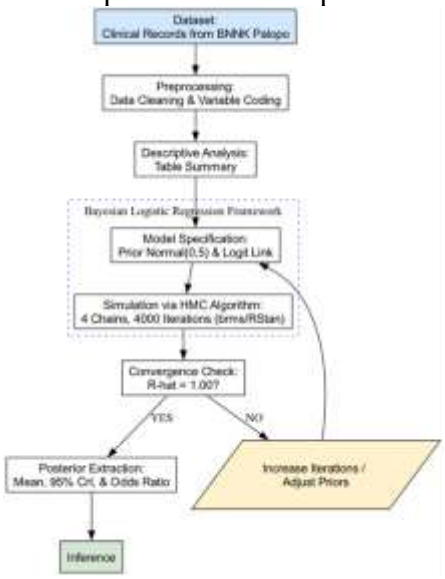


Figure 1. Flowchart

The primary data analysis technique used was Bayesian Logistic Regression. Let $y_i \in (0,1)$ represent the binary dependent variable (stimulant use) for the i-th individual, which follows a Bernoulli distribution: $y_i \sim Bernoulli(p_i)$ where p_i is the probability that the i-th participant is a stimulant user. This probability is modelled against a set of k predictor variables x, using the logit link function:

$$\text{logit}\, p_i = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \qquad (1)$$

This methodology was specifically selected over the traditional frequentist Maximum Likelihood Estimation (MLE) because of constraints in the dataset. The sample size was relatively small, and the dependent variable was highly imbalanced (approximately 74% of participants were stimulant users). The Bayesian approach with weakly informative prior distributions can regularise

the estimates. According to Bayes' theorem, the posterior distribution $P(\beta|y,X)$ is proportional to the product of the likelihood function $P(y|X,\beta)$ and the prior distribution $P(\beta)$:

$$(\beta|y,X) \propto P(y|X,\beta)\,P(\beta) \quad (2)$$

Specifically, weakly informative priors were specified for the regression coefficients to regularise the estimates $\beta_j \sim N(0,\sigma^2)$. This regularisation can help stabilise the estimation of the parameters, which are subsequently transformed into Odds Ratios (OR=e^{β_j}) using posterior probability distributions.

The analysis was conducted using R. The posterior distributions were simulated using the Hamiltonian Monte Carlo (HMC) algorithm via the brms and RStan packages. HMC was chosen for its superior efficiency in navigating complex probability spaces, avoiding the random-walk behaviour of traditional Markov Chain Monte Carlo (MCMC) methods. Convergence of the sampler was checked using MCMC diagnostics. The simulation was run across four independent chains, each with 4,000 iterations (including 1,000 warmup iterations). Model convergence—serving as the simulation's reliability metric—was confirmed using the potential scale reduction factor $(\hat{R})$, where a value of 1.00 across all parameters indicates that the chains have converged. Evidence for an association was judged by whether the 95% credible intervals (CrI) included 1, without formal significance testing.

C. Result and Discussion

The distribution of demographic and clinical characteristics among the 74 outpatients is summarised in Table 2.

Table 2. Demographic and Clinical Characteristics of Outpatients

Characteristic	0	1	p-value[2]
	N= 19[1]	N = 55[1]	
X1	23 (20,28)	28 (24, 36)	0.004
X2			0.2
0	7 (37%)	14 (25%)	
1	3 (16%)	3 (5.5%)	
2	8 (42%)	36 (65%)	
3	1 (5.3%)	2 (3.6%)	
X3	17 (15,23)	23 (20,32)	<0.001
X4			0.6
0	1 (5.3%)	5 (9.1%)	
1	11 (58%)	36 (65%)	
2	7 (37%)	14 (25%)	

[1]Median (Q1, Q3); n(%)

[2]Wilcoxon rank sum test; Fisher's exact test

Table 2 summarizes the characteristics of the study participants categorized by their substance use pattern (0 = non-stimulants; 1 = stimulants). The descriptive analysis revealed significant disparities in the age-related variables. The Current Age (X1) was significantly higher in the stimulant group (median = 28, IQR = 24–36) than in the non-stimulant group (median = 23, IQR = 20–28), with a p-value of 0.004. More notably, the Initiation Age (X3) showed a highly significant difference ($p<0.001$), where stimulant users initiated substance use at a later stage of life (Median = 23) than their counterparts (Median = 17).

In contrast, categorical variables such as Occupational Status (X2) and Environmental Influence (X4) did not show statistically significant differences between the two groups (p-values of 0.2 and

0.6, respectively). Although not statistically significant, 65% of stimulant users were self-employed, compared with 42% of non-stimulant users. These findings provide preliminary context for estimating the odds ratios of these factors with a Bayesian model.

1. MCMC Convergence Diagnostics

Before performing formal inference on the risk factors, the reliability of the Bayesian Logistic Regression model was verified using Markov Chain Monte Carlo (MCMC) convergence diagnostics. The Hamiltonian Monte Carlo (HMC) algorithm was run across four independent chains, each with 4,000 iterations. The visual evidence of this process is shown in Figure 2, which displays both the posterior density histograms and trace plots for all parameters.

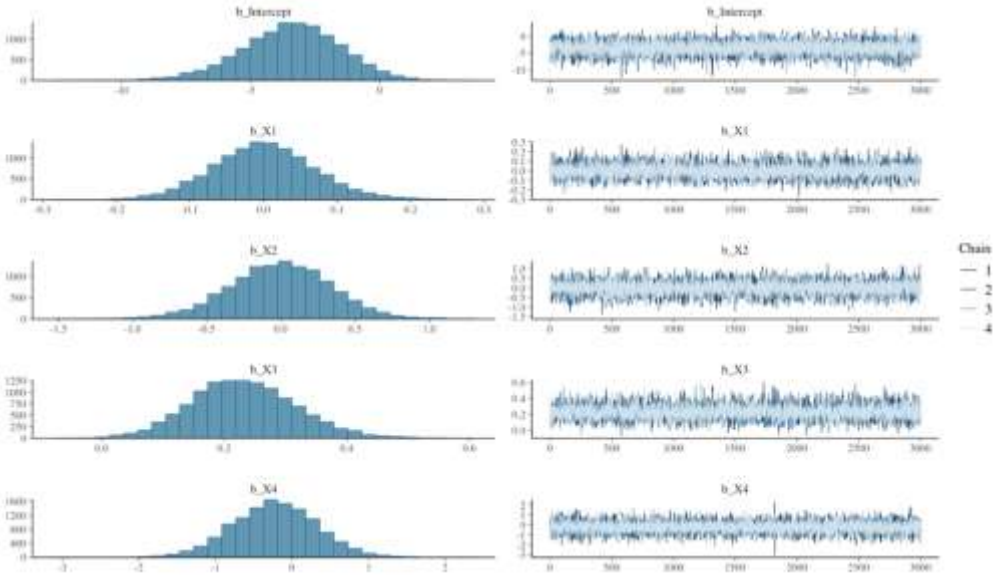


Figure 2. MCMC diagnostic plots

Visual inspection of the trace plots indicated that the chains were well mixed, suggesting that the algorithm successfully explored the target posterior distribution without significant autocorrelation. The potential scale reduction factor $(\hat{R})$ for all parameters—including the intercept and all independent variables—yielded a precise value of 1.00. This result confirmsindicates that the computationalchains process achieved full convergenceconverged. Despite the relatively small clinical sample size, the(n=74), estimatedthese valuesdiagnostics indicate that the generated posterior distributions are robust and validadequate for inferentialinference, analysisalthough the estimates remain imprecise.

2. Parameter Estimation and Posterior Odds Ratio

Bayesian logistic regression identified one factor associated with stimulant use among outpatients. Associations were interpreted using the 95% credible interval (CrI): an interval that excluded 1 was taken as support for an association, without formal significance testing. A summary of the parameter estimates and their corresponding posterior probabilities is presented in Table 3.

Table 3. Posterior Estimates and Odds Ratios for Stimulant Use

Var	OR (Estimate)	Est. Error	95% CI (Lower)	95% CI (Upper)
β	0.03	6.62	0.00	1.08
X1	1.00	1.07	0.88	1.15
X2	0.99	1.40	0.50	1.90
X3	1.26	1.09	1.08	1.51
X4	0.79	1.73	0.27	2.30

The Bayesian posterior estimates, presented as Odds Ratios in Table 3, show that Initiation Age (X3) was the only factor whose credible interval excluded 1. With a posterior median of 1.26 and a 95% credible interval of [1.08; 1.51], an older age at substance initiation was associated with higher odds of stimulant use. The credible intervals for current age, occupation, and environment included 1, so these data do not support an association for them.

The distribution of these risk factors is further illustrated in Figure 3, which shows the posterior densities of the risk factors. This study evaluated the influence of Current Age (X1), Occupational Status (X2), Initiation Age (X3), and Environmental Influence (X4) on the odds of stimulant use. Evidence for an association was judged by whether the 95% credible intervals (CrI) included the null value (0 for log-odds or 1 for odds ratios).

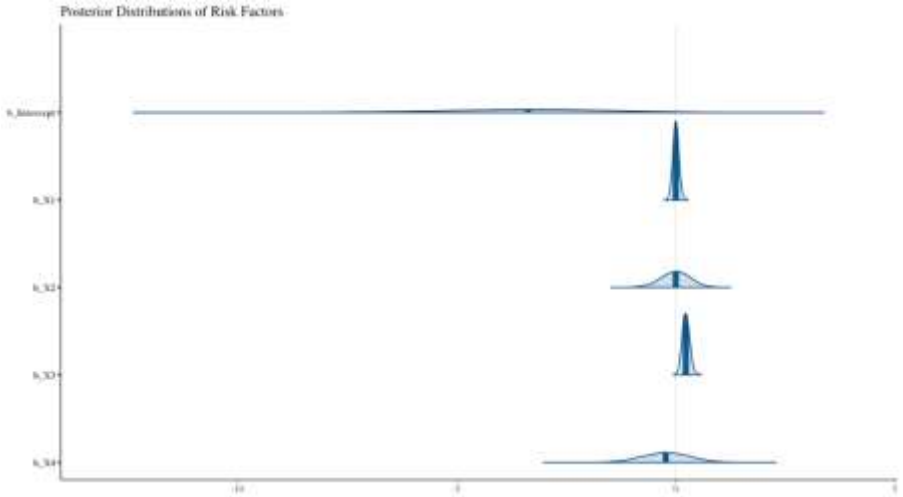


Figure 3. Posterior Distributions of the Analysed Factors

Figure 3 illustrates the posterior density distributions for the four analyzed factors (X1, X2, X3, and X4), where each ridge represents the probability distribution of the regression coefficients (β) updated using clinical data from BNNK Palopo. The vertical grey line at 0 marks no association on the log-odds scale. A variable was considered to have supporting evidence of an association if its 95% credible interval did not include zero. Notably, the distribution for Initiation Age (X3) stands out, as it is entirely to the right of the zero line, indicating a positive association with stimulant use; as the age of initiation increases, the odds of stimulant use also increase. In contrast, the distributions for Current Age (X1), occupation (X2), and environment (X4) all overlap with the zero-axis, indicating that no consistent association was detected for these factors in this outpatient sample. Furthermore, the varying widths of the distributions provide insight into estimation precision; the narrow peak for X3 indicates relatively higher precision, whereas the wider spread for X4 reflects greater uncertainty about environmental influences among the respondents.

3. Discussion

Descriptively, 55 of the 74 outpatients (74%) were coded as stimulant users, that is, methamphetamine (sabu-sabu) as the only recorded substance, and 13 of the 19 non-stimulant records involved methamphetamine combined with other substances. These figures describe only the outpatients analysed at BNNK Palopo and cannot be generalized to the wider population of drug users; broader epidemiological data are needed to determine the actual burden of stimulant use in the region. From a neurobiological perspective, the rapid dopaminergic surge produced by stimulants has been linked to strong reinforcement and may make treatment engagement and recovery more challenging. One of the main findings of this study is the association between initiation age and stimulant use. The Bayesian estimation yielded an Odds Ratio of 1.26, indicating that for every one-year increase in the age of first substance use, the odds of stimulant use increased by 26%. The literature mostly links early initiation to more severe dependence; these results concern a different outcome, namely whether a patient was coded as a stimulant user, and should not be read as contradicting that literature. Patients coded as non-stimulant, including those who combined methamphetamine with other substances, began substance use earlier (median 17 vs 23 years), which is consistent with the possibility that earlier initiation is linked to broader experimentation, although this study did not measure that mechanism. Because current age and initiation age were strongly correlated ($r \approx 0.81$), their separate contributions should be interpreted cautiously, and any implication for prevention is tentative.

Environmental influence (X4) had a posterior median odds ratio of 0.79, but its wide credible interval (0.27–2.30) included 1, so the data neither support nor rule out an association. Environmental influence was defined as a three-category nominal variable, whereas Table 3 reports a single coefficient for it; the estimate should therefore be interpreted cautiously until the variable is modelled as categorical. Previous studies suggest that the social environment can act as a trigger through peer pressure or as a buffer through social supervision, which warrants testing in larger samples.

Current age (X1) and occupation (X2) had posterior median odds ratios near 1.00 with credible intervals that included 1, so no association was detected after accounting for the other variables. This should not be read as evidence of no effect: the sample size was small, and current age was strongly correlated with initiation age ($r \approx 0.81$), which may have attenuated the association.

Methodologically, the sample was small (74 patients, 19 in the smaller outcome category) relative to the number of model parameters, so weakly informative priors were used to regularise the estimates. The HMC algorithm converged satisfactorily ($\hat{R}$=1.00). Convergence, however, does not guarantee that the model is correctly specified. Bayesian estimates depend on the prior, and with a sample of this size, the prior may have a non-negligible influence; a prior sensitivity analysis and a comparison with a penalised-likelihood approach would strengthen the conclusions. Other limitations are the single-site retrospective design, the absence of additional potential confounders, and a non-stimulant group that combines heterogeneous patterns of use, including methamphetamine used together with other substances.

D. Conclusion

This study, which applies Bayesian Logistic Regression analysis to outpatient data at BNNK Palopo, provides an exploratory estimate of the factors associated with stimulant use. Methodologically, the Hamiltonian Monte Carlo (HMC) algorithm converged satisfactorily (R-hat = 1.00 for all parameters); however, the estimates depend on the chosen priors and were not compared with frequentist alternatives in this study.

Substantively, initiation age (X3) was the only factor whose 95% credible interval excluded 1: with each one-year increase in age at first drug use, the odds of stimulant use (methamphetamine

only) rose by 26% (OR 1.26; 95% credible interval 1.08–1.51), whereas earlier initiation was associated with the non-stimulant group, which included polysubstance use. The credible intervals for environmental influence (OR 0.79; 0.27–2.30), current age, and occupational status included 1, so no association was detected for them. These findings are exploratory, come from a single site with 74 patients, and cannot be generalized beyond this outpatient sample.

Future studies should draw on broader epidemiological data to determine which substances pose the greatest burden in the BNNK Palopo working area (Palopo, Luwu, North Luwu, and East Luwu), because the present findings describe only the outpatients analyzed and cannot be generalized to the wider population of users. Such data could include district-level prevalence surveys, routine surveillance of treatment admissions by substance type, and laboratory or law enforcement records linked to the same period. We also recommend that BNNK Palopo record substance use patterns consistently at assessment, distinguishing single-substance from polysubstance use, and routinely capture age at first use, occupational status, and environmental influence, so that larger multi-period or multi-site analyses with appropriate categorical coding and prior sensitivity checks can be conducted.

Statement on the Use of Artificial Intelligence

- *The authors used Paperpal for language editing and Claude AI to assist with preliminary analysis. All outputs were were checked and verified by the authors, who take full responsibility for the content of the article.*

Author Contribution Statement

Penulis	*Contribution*
Besse Helmi Mustawinar	*Conceptualization, methodology, formal analysis, writing – original draft, writing – review and editing*
Yuliani	*Methodology, visualization, writing – review*
Dwi Risky Arifanti	*Data curation, investigation, validation of results*
Ikes Dwiastuti	*Investigation, field data collection, validation of results*
Tsabitah Fauziyah Azzahra	*Literature review, writing – original draft, editing*
Nasti Novyandini	*Project administration, writing – review and editing*

Funding Statement

This research did not receive funding from any institution, whether governmental, private, or non-profit.

Conflict of Interest Statement

The authors declare that there is no conflict of interest related to the research, authorship, or publication of this article.

References

Choi, N. G., Dinitto, D. M., Marti, C. N., & Choi, B. Y. (2021). U.S. older adults' heroin and psychostimulant use treatment admissions, 2012–2019: Sociodemographic and clinical characteristics. *Drug and Alcohol Dependence*, *231*, 109256. https://doi.org/10.1016/j.drugalcdep.2021.109256

Knopf, A. (2023a). NSDUH finds substance use, depression, suicidality high among youth. *The Brown University Child and Adolescent Behavior Letter*, *40*(1), 5–6. https://doi.org/10.1002/cbl.30759

Knopf, A. (2023b). NSDUH: More consumers reporting SUD, mental health problems. *Alcoholism & Drug Abuse Weekly*, *35*(3), 4–6. https://doi.org/10.1002/adaw.33663

Nandini, Andi Bunayya. 2025, Desember 31. BNN Palopo Rehabilitasi 233 Penyalahguna Narkotika Sepanjang 2025. Tribun Timur. https://makassar.tribunnews.com/palopo/1823789/bnn-palopo-rehabilitasi-233-penyalahguna-narkotika-sepanjang-2025.

Nath, A., Choudhari, S. G., Dakhode, S. U., Rannaware, A., & Gaidhane, A. M. (2022). Substance Abuse Amongst Adolescents: An Issue of Public Health Significance. *Cureus*, *14*(11). https://doi.org/10.7759/cureus.31193

Paul, K. I., Scholl, H., Moser, K., Zechmann, A., & Batinic, B. (2023). Employment status, psychological needs, and mental health: Meta-analytic findings concerning the latent deprivation model. *Frontiers in Psychology*, *14*(909558). https://doi.org/10.3389/fpsyg.2023.1017358

Pavlou, M., Omar, R. Z., & Ambler, G. (2024). Penalized Regression Methods With Modified Cross-Validation and Bootstrap Tuning Produce Better Prediction Models. *Biometrical Journal. Biometrische Zeitschrift*, *66*(5), e202300245. https://doi.org/10.1002/bimj.202300245

Pomrenze, M. B., Paliarin, F., & Maiya, R. (2022). Friend of the Devil: Negative Social Influences Driving Substance Use Disorders. *Frontiers in Behavioral Neuroscience*, *16*, 836996. https://doi.org/10.3389/fnbeh.2022.836996

Rahman, M. M., Rahaman, M. M., Hamadani, J. D., Mustafa, K., & Shariful Islam, S. M. (2016). Psycho-social factors associated with relapse to drug addiction in Bangladesh. *Journal of Substance Use*, *21*(6), 627–630. https://doi.org/10.3109/14659891.2015.1122099

Rioux, C., Castellanos-Ryan, N., Parent, S., Vitaro, F., Tremblay, R. E., & Séguin, J. R. (2018). Age of Cannabis Use Onset and Adult Drug Abuse Symptoms: A Prospective Study of Common Risk Factors and Indirect Effects. *The Canadian Journal of Psychiatry*, *63*(7), 457–464. https://doi.org/10.1177/0706743718760289

Rosner, B., Neicun, J., Yang, J. C., & Roman-Urrestarazu, A. (2020). Substance use among sexual minorities in the US – Linked to inequalities and unmet need for mental health treatment? Results from the National Survey on Drug Use and Health (NSDUH). *Journal of Psychiatric Research*, *135*, 107–118. https://doi.org/10.1016/j.jpsychires.2020.12.023

Roux, P., Faye, A., Sagaon-Teyssier, L., Donadille, C., Briand Madrid, L., Carrieri, M. P., Maradan, G., Jauffret-Roustide, M., Lalanne, L., & Auriacombe, M. (2024). Prevalence of stimulant use and the role of opioid agonist treatment among people who inject drugs in France: Results from the COSINUS cohort study. *Drug and Alcohol Review*, *44*(1). https://doi.org/10.1111/dar.13955

Sathe, N., Yu, F., Dai, L., Cribb, D., Spagnola, K., Saravia, A., Cai, R., & Hoenig, J. (2022). Quality Assessment of the 2014 to 2019 National Survey on Drug Use and Health (NSDUH) Public Use Files. In *Lecture Notes in Computer Science* (pp. 324–346). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-031-13945-1_23

Shoaa Kazemi, M., Mahamid, F., & Hamamra, B. (2025). Exploring psycho-social growth after addiction trauma in families freed from substance use. Journal of Ethnicity in Substance Abuse, 1–19. https://doi.org/10.1080/15332640.2025.2564105

Sinha, R., Easton, C., & Kemp, K. (2003). Substance Abuse Treatment Characteristics of Probation-Referred Young Adults in a Community-Based Outpatient Program. *The American Journal of Drug and Alcohol Abuse*, *29*(3), 585–597. https://doi.org/10.1081/ada-120023460

Sun, J., & Reddy, C. K. (2013). *Big data analytics for healthcare*. 1525. https://doi.org/10.1145/2487575.2506178

The National Survey on Drug Use and Health (NSDUH). (2003). American Psychological Association. https://doi.org/10.1037/e441772005-001

Uthis, P., Hafifah, I., Suktrakul, S., Sornlar, K., & Sawangpon, M. (2025). The Resilience on Amphetamine Relapse Youth Scale: development and psychometric properties. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-90294-8

Yamani, I. A. J., & Alkhatib, A. J. (2025). Substance Abuse and Societal Impact. *Journal of Posthumanism*, *5*(6), 4667–4685. https://doi.org/10.63332/joph.v5i6.2672

Zhao, Y., Tourais, J., Pierce, I., Nitsche, C., Treibel, T. A., WeingäRtner, S., Schweidtmann, A. M., & Tao, Q. (2024). Bayesian Uncertainty Estimation by Hamiltonian Monte Carlo: Applications to Cardiac MRI Segmentation. *Machine Learning for Biomedical Imaging*, *2*(June 2024), 856–887. https://doi.org/10.59275/j.melba.2024-88fa

Zhu, J., Pu, S., He, J., Su, D., Cai, W., Xu, X., & Liu, H. (2024). Processing imbalanced medical data at the data level with assisted-reproduction data as an example. *BioData Mining*, *17*(1), 29. https://doi.org/10.1186/s13040-024-00384-y

Zhu, Z., Fang, X., Zhu, C., Luo, W., & Zhu, Y. (2024). Structured Prediction in Latent Subspace for Unsupervised Fault Diagnosis With Small and Imbalanced Data. *IEEE Sensors Journal*, *24*(15), 25106–25115. https://doi.org/10.1109/jsen.2024.3415642