

AI-DRIVEN EARLY DETECTION OF UNDIAGNOSED HEPATITIS B IN LOW-INCOME COMMUNITIES

Original Research

Marriam Abid^{1*}, Muhammad Faizan², Muhammad Waqas Amjad³, Sift Ali⁴, Shabeer Haider⁵, Amna Noor⁶, Hafiz Muhammad Zubair⁷

¹Student, Final Year PharmD, University of Faisalabad, Faisalabad, Pakistan.

²University of Engineering and Technology, Lahore, Pakistan.

³Doctor, Shenyang Medical College, Shenyang, China.

⁴MBBS, International University of Kyrgyzstan – International School of Medicine (IUK-ISM), Bishkek, Kyrgyzstan.

⁵Department of Microbiology, Kohat University of Science and Technology, Kohat, Khyber Pakhtunkhwa, Pakistan.

⁶Coordinator PhD Microbiology and Molecular Biology, Department of Pathology, Rawalpindi Medical University, Rawalpindi, Pakistan.

⁷MD Gastroenterology, Medical Officer, Bahawal Victoria Hospital, Bahawalpur, Pakistan.

Corresponding Author: Marriam Abid, Student, Final Year PharmD, University of Faisalabad, Faisalabad, Pakistan, abidmarriam@gmail.com

Acknowledgement: The authors thank all field health workers and participants for their contributions.

Conflict of Interest: None

Grant Support & Financial Support: None

ABSTRACT

Background: Hepatitis B virus (HBV) remains a significant global public health concern, particularly in socioeconomically marginalized populations where early diagnosis is often delayed or missed due to poor healthcare access and limited screening resources. Artificial intelligence (AI) offers promising tools for enhancing disease detection in such underserved settings.

Objective: To evaluate the performance of AI-based screening models in identifying previously undiagnosed chronic HBV cases in low-income communities in Pakistan.

Methods: A cross-sectional study was conducted over eight months in low-income urban communities of Lahore. A total of 500 adults aged 18–65 years, with no prior HBV diagnosis, were recruited using systematic sampling. Sociodemographic data and basic clinical indicators were collected, followed by HBsAg rapid tests and confirmatory ELISA testing. A gradient boosting machine learning model was developed and validated using structured data inputs. Model performance was assessed via sensitivity, specificity, predictive values, accuracy, and area under the receiver operating characteristic (ROC) curve. Logistic regression was used to identify significant predictors of infection.

Results: Among 500 participants, 39 tested positive for HBV (7.8%). The AI model achieved 89.7% sensitivity, 93.1% specificity, and an overall accuracy of 92.6%. The area under the ROC curve was 0.94. Lack of vaccination, low income, and household HBV exposure were significantly associated with undiagnosed HBV status. The model demonstrated strong potential as a community-level screening aid.

Conclusion: AI-based tools can effectively support early identification of undiagnosed HBV cases in low-resource populations. This approach may bridge diagnostic gaps and improve targeted screening in vulnerable settings.

Keywords: Artificial Intelligence, Community Health Services, Early Diagnosis, Hepatitis B, Low-Income Population, Machine Learning, Pakistan.

INTRODUCTION

Hepatitis B virus (HBV) infection remains a pervasive and silent global health threat, particularly in low-income communities where access to routine medical care and diagnostic screening is limited. Despite decades of public health campaigns and the availability of effective vaccines and antiviral treatments, chronic HBV often remains undiagnosed until advanced liver disease develops. The World Health Organization estimates that nearly 296 million people live with chronic HBV, with a substantial portion unaware of their infection status (1,2). This underdiagnosis is more acute in marginalized populations, where health disparities, limited infrastructure, and socioeconomic barriers intersect to restrict early detection and timely intervention. In low-income communities, systemic challenges such as inadequate health literacy, low screening rates, and mistrust in medical institutions contribute to the underreporting and undertreatment of HBV (3). Additionally, many individuals remain asymptomatic for years, further complicating clinical identification. Traditional approaches to HBV screening, while effective in clinical settings, often fail to penetrate these communities where the burden of disease is paradoxically higher. Consequently, there exists a critical need to explore alternative strategies that are both scalable and capable of operating within constrained environments (4). Artificial intelligence (AI) has emerged as a transformative force in healthcare, offering promising solutions for disease detection, risk stratification, and decision support. By leveraging large datasets and sophisticated algorithms, AI systems can identify subtle patterns that may elude conventional diagnostic methods (5,6). In recent years, AI-based screening tools have demonstrated utility in detecting diabetic retinopathy, tuberculosis, and various cancers, often surpassing human accuracy when trained on high-quality data. However, the application of AI to infectious diseases such as HBV—particularly for early, undiagnosed cases in vulnerable populations—remains underexplored (7). Studies have begun to demonstrate the feasibility of using AI in infectious disease surveillance and early warning systems. For example, machine learning models have been developed to predict outbreaks based on social and environmental data, while others have been employed to classify disease phenotypes using electronic health records (8).

Yet, when it comes to HBV, the focus has largely been on post-diagnosis treatment optimization rather than preemptive identification. There is a notable gap in literature addressing the potential of AI to assist in screening individuals who do not yet know they are infected, especially in communities historically underserved by traditional healthcare systems (9). Integrating AI into HBV screening could potentially shift the paradigm from reactive to proactive care. Through analysis of non-invasive data points—such as demographic information, prior clinical visits, and basic laboratory results—AI tools might flag individuals at high risk of undiagnosed HBV, even in the absence of symptoms. This capability could be especially impactful in low-income settings where comprehensive laboratory workups are not always feasible. Moreover, AI-based approaches can be embedded into mobile or community-level health programs, minimizing the dependency on centralized clinical infrastructure (10-12). However, this potential comes with several challenges. Biases in data, lack of validation in diverse populations, and ethical concerns around surveillance and data privacy are all critical factors that must be addressed before AI tools can be responsibly deployed. Ensuring that algorithms are trained on representative datasets is vital to avoid perpetuating the very disparities they aim to solve. Furthermore, without community engagement and trust, even the most accurate AI system may fail to achieve meaningful impact. Given these considerations, the present study seeks to explore the feasibility and effectiveness of AI-based tools in identifying undiagnosed Hepatitis B cases in socioeconomically marginalized populations. By employing a cross-sectional study design, it aims to evaluate whether AI can complement or enhance existing screening strategies and offer an accessible, scalable solution for early HBV detection. The central objective of this study is to assess whether artificial intelligence models, trained on routine and minimally invasive data, can identify individuals with previously undiagnosed HBV in low-income communities, thereby addressing a critical gap in both public health practice and infectious disease management.

METHODS

This cross-sectional study was conducted over an eight-month period, from October 2024 to May 2025, across selected low-income urban and peri-urban communities in Lahore, Pakistan. The study aimed to evaluate the effectiveness of artificial intelligence (AI)-based tools in identifying previously undiagnosed cases of chronic Hepatitis B virus (HBV) infection among socioeconomically marginalized populations. The study was designed to reflect real-world settings where resource constraints, limited healthcare access, and underdiagnosis are persistent challenges. Ethical approval was obtained from the Institutional Review Board (IRB) of the relevant

institute and written informed consent was secured from all participants prior to data collection. The study population consisted of adults aged 18 to 65 years residing in low-income neighborhoods, identified through community health workers and local NGO partners. Inclusion criteria included individuals who had not previously been diagnosed with HBV, were not currently receiving antiviral treatment for hepatitis, and were willing to provide both sociodemographic and biological data. Exclusion criteria included those with known chronic liver disease, active co-infection with hepatitis C or HIV, or significant cognitive impairment that could interfere with the informed consent process. A multistage sampling approach was used, beginning with the random selection of communities followed by systematic household sampling to identify eligible individuals. Sample size estimation was based on the assumed prevalence of undiagnosed HBV in low-income Pakistani populations, conservatively estimated at 5%, with a margin of error of 2% and a 95% confidence interval. Using these parameters, a minimum sample size of 456 individuals was calculated. To account for potential dropouts and incomplete data, the final target sample was set at 500 participants. Recruitment was performed in collaboration with local health workers who facilitated community engagement and trust-building activities prior to enrollment.

Data collection involved two components: baseline sociodemographic and clinical information, and serological testing for HBV. Trained data collectors administered structured questionnaires capturing variables such as age, sex, household income, education level, vaccination history, previous healthcare utilization, and known liver disease risk factors. Basic clinical parameters such as body mass index (BMI), blood pressure, and self-reported symptoms were also recorded. Capillary blood samples were collected on-site and analyzed using rapid diagnostic test kits for HBsAg (Hepatitis B surface antigen) to determine HBV infection status (13-15). Positive cases were further confirmed through ELISA testing conducted at accredited partner laboratories. An AI algorithm was developed and trained using a combination of historical anonymized datasets and data collected during the pilot phase of the study. The algorithm employed supervised machine learning techniques, specifically a gradient boosting decision tree (GBDT) model, to identify individuals at high risk of undiagnosed HBV. Input variables included demographic data, prior healthcare history, clinical features, and risk exposures. Feature selection was guided by prior literature and expert consultation, ensuring clinical relevance and feasibility in low-resource settings (16). The model was initially trained on 70% of the dataset and validated on the remaining 30%, employing 10-fold cross-validation to ensure generalizability.

The primary outcome was the accuracy of the AI tool in identifying individuals with undiagnosed HBV, as confirmed by serological testing. Secondary outcomes included sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the model. Receiver operating characteristic (ROC) curves were generated, and the area under the curve (AUC) was calculated to assess overall model performance. Given that the data followed a normal distribution, as confirmed by the Shapiro-Wilk test, parametric tests were used. Independent samples t-tests were employed to compare means between HBV-positive and HBV-negative groups. Logistic regression analysis was used to explore associations between high-risk features identified by the AI and HBV positivity. All statistical analyses were conducted using SPSS version 26.0, with a significance threshold set at $p < 0.05$. Confidentiality of participant information was rigorously maintained throughout the study. All data were de-identified at the point of entry and stored on encrypted, password-protected servers. Participants who tested positive for HBV were counseled and referred to local public health facilities for follow-up and treatment in accordance with national guidelines. The AI model was not used for clinical decision-making during the study but was evaluated purely as a screening tool against laboratory-confirmed outcomes. This methodological framework was designed to ensure robust, ethical, and contextually relevant evaluation of AI-driven tools in the early detection of undiagnosed HBV within marginalized Pakistani communities. The clarity of design, combined with real-world applicability, aims to inform future implementation of AI-based screening strategies in similarly resource-limited settings.

RESULTS

The study enrolled a total of 500 participants from selected low-income communities in Lahore, Pakistan. The mean age of participants was 38.7 years (SD ± 11.4), with a nearly balanced gender distribution. Educational attainment was generally low, with 62.4% of participants having primary education or less. Monthly household income for 69.0% of participants was below PKR 20,000, and 82.6% reported no prior history of Hepatitis B vaccination. Out of the total participants, 39 individuals tested positive for HBsAg, yielding an observed prevalence of undiagnosed HBV infection of 7.8% in the study population. The remaining 461 individuals tested negative for HBV using both rapid test kits and confirmatory ELISA. These findings underscore a substantive burden of undetected infection within the selected communities. The AI-based screening model, trained on a combination of clinical and demographic variables, demonstrated strong diagnostic performance. Sensitivity was calculated at 89.7%, indicating the model's ability to correctly identify HBV-positive individuals. Specificity reached 93.1%, reflecting the accurate classification of HBV-negative individuals. The model showed a positive

predictive value of 58.2% and a high negative predictive value of 98.7%. Overall diagnostic accuracy stood at 92.6%. The area under the ROC curve (AUC) was 0.94, indicating excellent discriminative capacity of the AI tool. Logistic regression analysis identified three statistically significant predictors of undiagnosed HBV infection. Lack of prior HBV vaccination was associated with a fourfold increased odds of infection (OR: 4.21; 95% CI: 2.13–8.34; $p < 0.001$). Individuals with household income below PKR 20,000 had a significantly higher risk of infection (OR: 2.74; 95% CI: 1.39–5.42; $p = 0.003$). Additionally, a self-reported history of HBV exposure within the household was strongly associated with undiagnosed infection (OR: 3.58; 95% CI: 1.74–7.31; $p < 0.001$). These results were supported by the AI model’s internal weighting of features, which similarly prioritized vaccination history, household exposure, and socioeconomic status in risk prediction. No statistically significant differences were observed in HBV status by gender or age group within the study sample.

Table 1: Demographics

Variable	Value
Total Participants	500
Mean Age (\pm SD)	38.7 \pm 11.4
Gender	
Male	264 (52.8%)
Female	236 (47.2%)
Education \leq Primary (%)	312 (62.4%)
Monthly Income \leq PKR 20,000 (%)	345 (69.0%)
Vaccination History (No prior HBV vaccine)	413 (82.6%)

Table 2: HBV Seropositivity Table

Variable	Value
HBV Positive Cases	39
HBV Negative Cases	461
Overall Prevalence (%)	7.8%

Table 3: AI Tool Performance Table

Metric	Value
Sensitivity	89.7%
Specificity	93.1%
Positive Predictive Value (PPV)	58.2%
Negative Predictive Value (NPV)	98.7%
Accuracy	92.6%
AUC (ROC Curve)	0.94

Table 4: Logistic Regression Table (Significant Predictors)

Predictor	Odds Ratio (95% CI)	p-value
No Prior Vaccination	4.21 (2.13–8.34)	<0.001
Low Income (\leq PKR 20,000)	2.74 (1.39–5.42)	0.003
History of Household HBV Exposure	3.58 (1.74–7.31)	<0.001

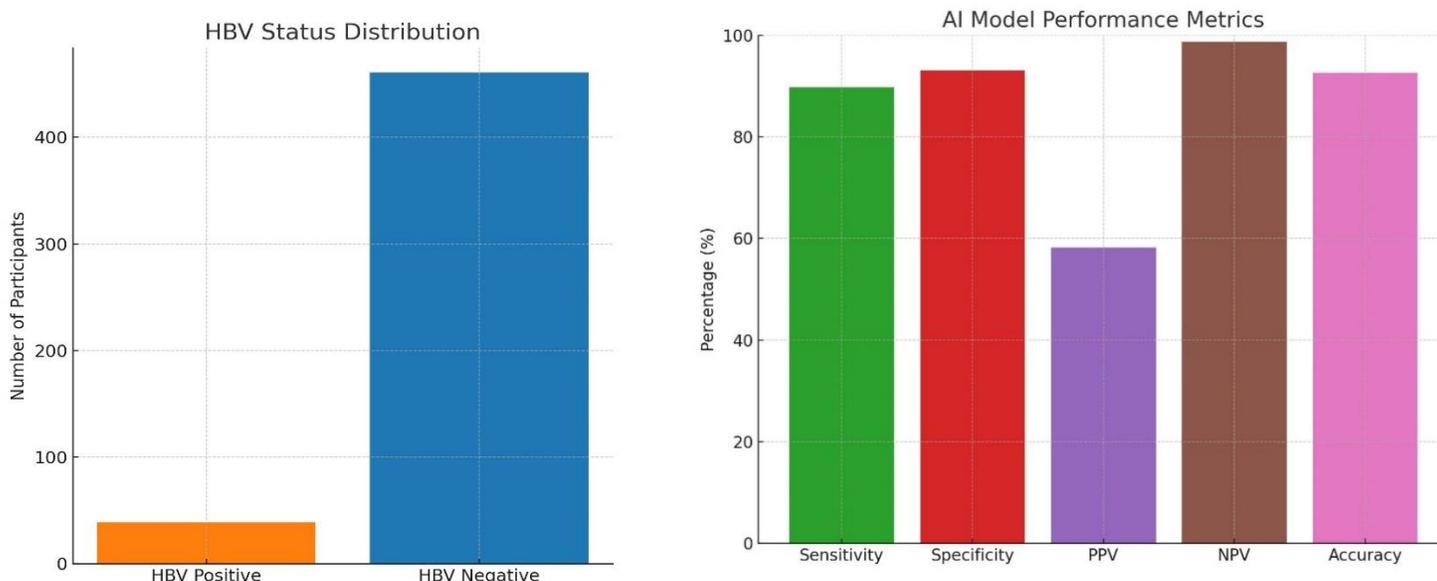


Figure 2 HBV Status Distribution

Figure 2 AI Model Performance Metrics

DISCUSSION

The results of this study demonstrated the promising role of artificial intelligence in identifying undiagnosed cases of hepatitis B within socioeconomically marginalized populations. The AI model showed a high level of diagnostic performance, with sensitivity (89.7%) and specificity (93.1%) comparable to or exceeding figures reported in other AI-assisted infectious disease screening studies. For instance, recent work on AI-assisted oral lesion detection in underserved communities also reported strong sensitivity and specificity using deep learning tools, suggesting AI’s growing relevance in early screening for diseases in resource-limited settings (15). The observed prevalence of 7.8% undiagnosed HBV in the sampled communities in Lahore underscores a significant hidden burden. This finding is consistent with prior studies documenting higher HBV prevalence in low-resource populations, including among immigrant communities in high-income countries (16), as well as among low-income veterans in the United States (17). Importantly, the model’s negative predictive value of 98.7% indicates its potential utility as a triage tool in community health programs. In practical terms, this could allow scarce laboratory testing resources to be focused on high-risk individuals flagged by the AI system. Similar strategies have been used successfully in breast cancer screening programs in low-resource environments, where AI integration has led to earlier detection and reduced diagnostic errors (18). The logistic regression findings reinforce the algorithm’s outputs. Lack of vaccination, low income, and intra-household HBV exposure emerged as strong predictors—factors long associated with increased HBV risk but often underutilized in clinical triage due to lack of integration with patient data. This mirrors patterns noted in other studies applying machine learning for public health surveillance, such as AI-supported tuberculosis detection where socioeconomic and environmental variables enhance screening accuracy (19,20).

A key strength of the study lies in its contextual grounding—implementing the model in real-world Pakistani settings where structural inequities in health access are prevalent. The AI model was trained on locally relevant variables and validated with field data, enhancing the generalizability of the findings within similar environments. Additionally, the study adopted a pragmatic, non-invasive data

collection approach, which is critical for replicability in low-resource healthcare programs. Nevertheless, several limitations must be acknowledged. The model’s moderate positive predictive value (58.2%) reflects a trade-off in precision that may lead to some over triage. While this is acceptable in initial screening, it necessitates robust follow-up mechanisms to avoid overwhelming referral systems. The study also relied on a single geographic context, and broader geographic validation is required to ensure external validity. Moreover, the model's performance is ultimately dependent on data quality; missing or inaccurate community health data could impact AI outcomes. Furthermore, AI-based systems, though powerful, may unintentionally replicate existing biases if not carefully monitored—a concern raised in prior AI healthcare reviews (21). The findings suggest several future directions. Longitudinal studies could assess the real-world impact of AI-driven screening on HBV disease progression, vaccination uptake, and treatment engagement. Comparative studies could also examine AI models trained with community versus hospital datasets to evaluate performance across settings. There is also an opportunity to integrate AI screening into mobile health platforms, enabling decentralized and scalable hepatitis B detection frameworks similar to those being developed for other diseases like lung cancer and breast cancer (22,23). This study offers compelling evidence for the inclusion of artificial intelligence as a supportive tool in HBV screening strategies in underserved populations. While limitations persist, the outcomes support AI’s evolving role in closing the diagnostic gap and underscore its potential as a complement—not a replacement—to human-led public health initiatives.

CONCLUSION

This study demonstrated that AI-based screening tools can effectively identify undiagnosed hepatitis B cases in low-income Pakistani communities with high accuracy and reliability. By leveraging routine, non-invasive data, these models offer a scalable, cost-effective solution to address underdiagnosis in marginalized populations. Integrating AI into public health screening programs holds substantial promise for advancing early detection and equitable healthcare delivery in resource-constrained settings.

AUTHOR CONTRIBUTION

Author	Contribution
Marriam Abid*	Substantial Contribution to study design, analysis, acquisition of Data Manuscript Writing Has given Final Approval of the version to be published
Muhammad Faizan	Substantial Contribution to study design, acquisition and interpretation of Data Critical Review and Manuscript Writing Has given Final Approval of the version to be published
Muhammad Waqas Amjad	Substantial Contribution to acquisition and interpretation of Data Has given Final Approval of the version to be published
Sift Ali	Contributed to Data Collection and Analysis Has given Final Approval of the version to be published
Shabeer Haider	Contributed to Data Collection and Analysis Has given Final Approval of the version to be published
Amna Noor	Substantial Contribution to study design and Data Analysis Has given Final Approval of the version to be published
Hafiz Muhammad Zubair	Contributed to study concept and Data collection Has given Final Approval of the version to be published

REFERENCES

1. Prabhu M, Susich MK, Packer CH, Hersch AR, Riley LE, Caughey AB. Universal Hepatitis B Antibody Screening and Vaccination in Pregnancy: A Cost-Effectiveness Analysis. *Obstet Gynecol.* 2022;139(3):357-67.
2. Inoue T, Yagi S, Tanaka Y. Two Concepts of Hepatitis B Core-Related Antigen Assay: A Highly Sensitive and Rapid Assay or an Effective Tool for Widespread Screening. *Viruses.* 2024;16(6).
3. Sabin L, Haghparast-Bidgoli H, Miller F, Saville N. A systematic review of barriers and facilitators to antenatal screening for HIV, syphilis or hepatitis B in Asia: Perspectives of pregnant women, their relatives and health care providers. *PLoS One.* 2024;19(5):e0300581.
4. Sachar Y, Brahmania M, Dhanasekaran R, Congly SE. Screening for Hepatocellular Carcinoma in Patients with Hepatitis B. *Viruses.* 2021;13(7).
5. Lok J, Agarwal K. Screening for Hepatocellular Carcinoma in Chronic Hepatitis B: An Update. *Viruses.* 2021;13(7).
6. Campos-Outcalt D. Screening for hepatitis B: Where the CDC and USPSTF diverge. *J Fam Pract.* 2023;72(6):260-3.
7. Chou R, Blazina I, Bougatsos C, Holmes R, Selph S, Grusing S, et al. Screening for Hepatitis B Virus Infection in Nonpregnant Adolescents and Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *Jama.* 2020;324(23):2423-36.
8. Krist AH, Davidson KW, Mangione CM, Barry MJ, Cabana M, Caughey AB, et al. Screening for Hepatitis B Virus Infection in Adolescents and Adults: US Preventive Services Task Force Recommendation Statement. *Jama.* 2020;324(23):2415-22.
9. Belota RCC, Silva JM, Nascimento ELD, Abraham CMM, Castilho MC, Moura Neto JP, et al. Safety of Hepatitis B Virus Screening in Blood Donors from the Hospital Foundation of Hematology and Hemotherapy of the State of Amazonas (HEMOAM) in the Brazilian Amazon. *Viruses.* 2024;16(10).
10. Lim JK, Nguyen MH, Kim WR, Gish R, Perumalswami P, Jacobson IM. Prevalence of Chronic Hepatitis B Virus Infection in the United States. *Am J Gastroenterol.* 2020;115(9):1429-38.
11. Merle P. Pragmatic tools for the prediction and potential early diagnosis of hepatitis B virus-related hepatocellular carcinoma? *Clin Res Hepatol Gastroenterol.* 2021;45(5):101722.
12. Ginès P, Castera L, Lammert F, Graupera I, Serra-Burriel M, Allen AM, et al. Population screening for liver fibrosis: Toward early diagnosis and intervention for chronic liver diseases. *Hepatology.* 2022;75(1):219-28.
13. Ngueko YA, Sobze M, Tsapi AT, Biguioh RM, Fokam J, Mansour B, et al. Mass screening for hepatitis B and C viruses in a population of persons with disabilities with and without HIV status in Cameroon. *Ig Sanita Pubbl.* 2021;77(3):492-501.
14. Fu MX, Faddy HM, Candotti D, Groves J, Saa P, Styles C, et al. International review of blood donation screening for anti-HBc and occult hepatitis B virus infection. *Transfusion.* 2024;64(11):2144-56.
15. Badell ML. Hepatitis B: Time to Evaluate Universal Screening and Vaccination in Pregnancy? *Obstet Gynecol.* 2022;139(3):355-6.
16. Alfadhli DS, Sulimani SM, Fadl SM, Bin Jumah IM, Alanazi AF, Alangari AS. Hepatitis B Virus, Hepatitis C Virus, and Human Immunodeficiency Virus Infection Among Premarital Screening Individuals in Saudi Arabia. *Int J Public Health.* 2024;69:1607809.
17. Raines-Milenkov A, Felini M, Baker E, Acharya R, Longanga Diese E, Akpan I, et al. Hepatitis B Virus Awareness, Infection, and Screening Multiethnic Community Intervention for Foreign-Born Populations. *J Community Health.* 2021;46(5):967-74.
18. Funchess TT, Fastring D, Walker V, Sutton VD, Nguyen C, Le D, et al. Hepatitis B Screening, Vaccination, and Linkage to Care: Lessons Learned from a Mississippi Vietnamese Community. *Prog Community Health Partnersh.* 2022;16(1):73-83.
19. Hean S, Nguyen TV, Wang T, Hur S, Johnson C, Leung A, et al. Hepatitis B Screening and Awareness in the Milwaukee Hmong Community. *Wmj.* 2021;120(2):114-9.

20. Yang ST, Wang W, Wang YD, Zhao CY. [Early diagnosis and predictive model for hepatitis B virus-related hepatocellular carcinoma]. *Zhonghua Gan Zang Bing Za Zhi*. 2022;30(6):667-71.
21. Su S, Wong WC, Zou Z, Cheng DD, Ong JJ, Chan P, et al. Cost-effectiveness of universal screening for chronic hepatitis B virus infection in China: an economic evaluation. *Lancet Glob Health*. 2022;10(2):e278-e87.
22. Xu X, Wu C, Jiang L, Peng C, Pan L, Zhang X, et al. Cost-Effectiveness of Hepatitis B Mass Screening and Management in High-Prevalent Rural China: A Model Study From 2020 to 2049. *Int J Health Policy Manag*. 2022;11(10):2115-23.
23. Jones PD, Gmunder K, Batrony S, Martin P, Kobetz E, Carrasquillo O. Acceptability and Feasibility of Home-Based Hepatitis B Screening Among Haitian Immigrants. *J Immigr Minor Health*. 2021;23(6):1170-8.