**Datathon 2023**

**Department of Statistics and Actuarial Science**

**The University of Iowa**

**Strategic Screening: Developing a Predictive Model for Chronic Hepatitis B Infection Using Reliable Patient Data**

Chronic Hepatitis B virus (HBV) infection can lead to cirrhosis, liver cancer, and liver failure. While treatable when detected early, many patients with chronic HBV infection remain asymptomatic and undiagnosed until severe liver damage occurs. By the time of diagnosis, the opportunity for effective treatment may have passed.

The implementation of HBV vaccination for U.S. children in 1991 has been highly effective in preventing the infection. However, despite the availability of these vaccines, there has been a stagnation in the reduction of HBV infections, and hepatitis-related deaths have risen. This may be partially attributable to immigration from regions with low vaccination rates or to instances where parents opt out of vaccinating their children.

One objective of the U.S. National Viral Hepatitis Action Plan is to reduce mortality and improve health outcomes for individuals with viral hepatitis. The Action Plan states that expediting the diagnosis, care, treatment, and potential cure of those with chronic viral hepatitis is crucial for reducing associated mortality and enhancing patient health.

Blood tests can determine a patient's history of HBV immunity and infection. The presence of antibodies to the hepatitis B core antigen (anti-HBc) signifies a past HBV infection. Current chronic infection is indicated by the hepatitis B surface antigen (HBsAg), while the hepatitis B surface antibody (anti-HBs) demonstrates immunity acquired through vaccination.

To achieve the Action Plan's objective, an extensive effort to collect and screen blood samples for chronic HBV infection could be undertaken, but this method would be costly. A more efficient strategy would be to pinpoint individuals at elevated risk for chronic HBV infection using readily available demographic and medical data, allowing health care providers to direct these high-risk individuals towards blood testing.

Use NHANES to develop this predictive model for chronic HBV infection risk. In a clinical context, such a model would primarily serve patients who are either unvaccinated or unsure of their vaccination status. Additionally, the model might exclude patients who have been previously diagnosed with an HBV infection by a health care professional. Can your estimation accurately reflect these specific populations?

When developing a predictive model for chronic HBV infection, it is critical to consider the reliability and practicality of the variables used. While certain behaviors such as illegal drug use and some sexually transmitted infections are associated with a higher risk of HBV infection, patients may not always disclose this information to health care providers, leading to underreporting and data inaccuracies.

Similarly, socioeconomic factors like income or net worth could predict HBV infection, yet patients may not know or wish to share precise financial details. Given these challenges, it is essential to identify which variables are both reliable and feasible for health care providers to collect during a typical patient encounter like a doctor’s office visit.

The model must be designed with sensitivity to these factors and incorporate data that can be reliably obtained during a medical visit. The variables chosen should be those that can be systematically and ethically collected, ensuring that the predictive model can be applied broadly and efficiently in a clinical setting.