COMPARING ESTIMATES OF HEV SEROPREVALENCE FROM HOSPITAL-BASED SURVEILLANCE WITH POPULATION-BASED SURVEYS

by
Sophie Suh-young Kang

A research thesis submitted to Johns Hopkins University
in conformity with the requirements for
the degree of Master of Science

Baltimore, Maryland
April 2020
Abstract

Hepatitis E (HEV) is an endemic infection with serious consequences in pregnant women that is not systematically surveilled for in Bangladesh. Serologic surveys measure the proportion of people in a population susceptible to a pathogen, which can reveal population level disease dynamics. This study investigates whether HEV seroprevalence estimates derived from hospital-based surveillance approximate those derived from a nationally representative population survey. We compared age, sex, and region-specific IgG seroprevalence estimates and characterized how subgroup seroprevalence estimates from the hospital-based surveillance are related to those from a national survey. We compared estimates from the two data sources using a Pearson’s chi-squared test adjusted with Rao and Scott’s first order correction for survey design. We also adjusted hospital division estimates by age and sex. The hospital survey had an overall seroprevalence estimate of 30% (95% CI: 28-33), which was not significantly different from that of the national survey, 25% (95% CI: 21-30). Male seroprevalence estimates were not statistically different between the two surveys ($\chi^2 = 1.85, p \text{ value} = 0.17$), with males in the national survey at 31% seropositivity (95% CI: 26%-35%) and 35% (95% CI: 26%-35%) in the hospital survey. Female estimates were also not significantly different ($\chi^2 = 0.08, p \text{ value} = 0.77$), estimated at 20% (95% CI: 16%-25%) in the national survey and 21% (95% CI: 18%-25%) in the hospital survey. No age group specific seroprevalence estimates were significantly different, except for the 30-39 year old age group ($\chi^2 = 5.11, p \text{ value} = 0.03$). In the Barisal division, the national survey population had an estimate of 47% (95% CI: 37%-57%) while the hospital survey’s estimate was 21% (95% CI: 18%-25%) which was a significant difference ($\chi^2 = 4.3, p \text{ value} = 0.03$). The other four out of the five division specific estimates had no significant difference. Seroprevalence of IgG antibodies to HEV estimated from a hospital-based convenience sample provided similar estimates to those generated from a systematic random sample of the general population in Bangladesh, suggesting that the less resource intensive hospital sampling approach is able to reflect the same inferences as a nationally representative sample. This strategy for surveillance to understand population-level disease susceptibility should be considered for different settings and infections where a nationally representative survey is not possible.
Table of Contents

I. Abstract...........................................................................................................................................ii

II. Introduction.....................................................................................................................................1

III. Methods.........................................................................................................................................3

   a. Nationally Representative Population Survey
   b. Hospital-based Surveillance
   c. Comparing Seroprevalence between the two studies
   d. Sensitivity Analysis

IV. Results ...........................................................................................................................................6

V. Discussion.......................................................................................................................................8

VI. Tables and Figures.........................................................................................................................14

VII. References....................................................................................................................................22

VIII. Curriculum Vitae.........................................................................................................................27
Introduction

Hepatitis E (HEV) is a common infection in resource-poor countries that is predominantly spread by the fecal-oral route through drinking water and undercooked meat\(^1\). The primary symptoms of HEV are inflammation of the liver and acute jaundice\(^1\). HEV has four genotypes, with genotype 1 and 2 mostly associated with causing waterborne outbreaks in developing economies and genotype 3 and 4 associated with exposure to wild animals particularly through the consumption of game meat\(^2\).

Though HEV is generally a self-limiting, acute infection with a case fatality rate lower than 1% in the greater population, special populations such as pregnant women and their fetuses can have serious adverse health outcomes\(^3\). Pregnant women with acute HEV infection have a much higher risk of poor birth outcomes and maternal mortality. The case fatality rate for HEV-infected pregnant women ranges from 10-25%, with one hospital study in Sudan reporting a rate as high as 28%\(^4,5\).

Despite the possibility of such serious outcomes, the prevalence and distribution of HEV infection are not well understood because of the lack of systematic surveillance and the large proportion of mild or asymptomatic cases that go undiagnosed because they do not require clinical care\(^2\). Even with symptomatic infection, HEV presents with symptoms that are similar to other hepatitis infections and requires a laboratory test for diagnosis, which contributes to the underdiagnosis of cases. Furthermore, low-income countries rarely have systemic surveillance in place for HEV. Even among high-income countries such as the European Union Member States, only 20 out of the 30 Member States have well established HEV surveillance\(^6\). These aspects of HEV make epidemiological studies difficult and are a major barrier to understanding the epidemiology and burden of the disease\(^2\).

Given these challenges, a serological study can be the most direct method to estimate population level disease susceptibility and the force of infection within the population\(^7\). Serologic tests look for specific antibodies to a pathogen in the blood or serum to indicate past or current infection with the virus. When one is infected with HEV, IgM antibodies appear during the early acute phase of infection and can be detectable for up to five months\(^8\). On the other hand, IgG antibodies remain in the serum for much longer, first
appearing during the acute phase of infection and remaining detectable for 3-14 years after infection; precisely how long IgG anti-HEV remains detectable is not fully agreed on\textsuperscript{8,9,10}. Detecting IgM anti-HEV in a serology tests indicates current infection while detecting IgG anti-HEV alone indicates past infection.

Serologic surveys measure the proportion of people in a population positive for a specific antibody to a pathogen, which can reveal population level disease dynamics\textsuperscript{7}. A nationally representative serological survey is the gold standard for making confident estimates of population susceptibility\textsuperscript{11}. In a disease like HEV with large proportions of subclinical infections, nationally representative serological studies are able reveal insights on the dynamics of infection within a population.

Nationally representative studies can be considered relatively resource intensive and logistically challenging, particularly during the sample collection stage\textsuperscript{7,12}. In contrast, hospital-based surveillance systems are common, even in low-income countries, routinely collect serum samples from patients for routine clinical care\textsuperscript{13,14}. Hospital-based surveillance systems capture a convenience sample of the population—people who seek care at a hospital—knowing that these populations can be significantly different from the overall population. Hospital surveillance systems are by definition syndromic, meaning that certain symptoms or a set of symptoms are reportable in a hospital system. Naturally, people who seek care at a hospital are generally have symptoms severe enough to warrant care-seeking. People who seek care at a hospital are also able to overcome any logistical barriers such as transportation, language, and cost to actually utilize the hospital\textsuperscript{15}. In Bangladesh, these barriers can result in differential hospital utilization based on religion, age, and sex\textsuperscript{16}.

An existing serum bank, collected from consenting patients presenting to hospitals, could be a valuable resource for serostudies but is limited from making population inferences by the likelihood of differences in a hospital sample from the greater population\textsuperscript{17}. That said, studies that actually demonstrate and describe immunological differences between hospital-based and nationally representative surveillance are scarce in the literature.

In Bangladesh, HEV is endemic with annual outbreaks being reported throughout the country and is thought to be a major cause of acute jaundice in adults. Until recently
nation-wide estimates of HEV infection and distribution were unavailable; this was fortunately amended by a nationally representative seroprevalence study conducted between 2014 and 2016\textsuperscript{16}. In the same time period, a hospital-based surveillance system collected data on patients presenting with acute jaundice in six tertiary hospitals. Blood
and serum samples were collected from all consenting patients with acute jaundice, regardless of diagnosis. The overlap of these two serological studies makes it possible to reasonably compare the different serostudy sampling methods and explore estimates of seroprevalence from a hospital-based serostudy is representitive of a nationally representative one.

This study investigates whether HEV seroprevalence estimates derived from bio-banked serum samples from hospital-based surveillance approximate those derived from a nationwide sero-survey. This comparison considers age, sex, and region-specific IgG seroprevalence estimates and will aim to characterize how subgroup seroprevalence estimates derived from the hospital-based surveillance are related to those from a national survey.

**Methods**

This is a secondary data analysis of one seroprevalence study and one surveillance dataset. The first study is a nationally representative population survey that was carried out between October 2015 and January 2016. The second study is a hospital-based surveillance study collecting data from six hospitals between December 2014 and September 2017\textsuperscript{16}. Both studies were led by the International Centre for Diarrheal Disease Research, Bangladesh (icddr,b), in conjunction with the Institute of Epidemiology, Disease Control, and Research, a research institute under the Bangladesh Ministry of Health.

**Nationally Representative Population Survey**

The nationally representative sero-study drew its population from individuals living in 70 randomly selected communities in Bangladesh\textsuperscript{11}. The study population was sampled using a two-stage cluster sample design. The primary sampling units were the communities, of which there were 70 communities were selected out of 97,162. The probability of a community being selected was proportional to the population size of the
community. From within those 70 communities, 707 households, the secondary sampling unit, were selected.

All household members over the age of 6 months were eligible. Blood samples were collected from all healthy individuals who consented. Participants were also asked to respond to a survey. Serum samples were tested for the presence of anti-HEV IgG antibodies using the Wantai immunoassay kit (Wantai HEV IgG [WE-7296] ELISA kit; Wantai Biological, Beijing, China).

Hospital-based Surveillance

The hospital-based acute jaundice surveillance system drew its population from acute jaundice patients at six tertiary hospitals in Bangladesh (Figure 1). The hospitals are regionally dispersed around Bangladesh, located in five of the seven divisions. Any patient who was hospitalized and presenting with acute jaundice was listed, and all those individuals over the age of 14 were eligible to be included in the study. Acute jaundice was defined as new onset of yellow eyes or skin within the last three months that continued on the day of admission. Written informed consent for enrollment in the study was obtained from patients over 17 years old. Written assent was obtained from patients between the ages of 14 and 17 before also obtaining written consent from their parents or guardians. Serum samples were tested for anti-HEV IgM and IgG antibodies using the same Wantai ELISA kits as the nationally representative serosurvey.

As our study is investigating the background seroprevalence among the hospital population, we considered only patients without active HEV infection, as identified through their anti-HEV IgM levels. Patients who were positive for anti-HEV IgM antibodies were excluded from the study to ensure that we were not over selecting participants who were seeking care for active HEV infection.

Comparing Seroprevalence between the two studies

We compared estimates of anti-HEV IgG antibodies prevalence derived from a hospital-based serosurvey with those derived from a nationally representative serosurvey. First, we found the overall seroprevalence in both populations, then stratified on three basic demographic characteristics: age, sex, and residential location.
We found the overall sex-specific seroprevalence estimates for both populations. We binned ages into ten-year age groups and found overall age-group seroprevalence estimates as well as sex-specific age group seroprevalence. Residential location was given at the level of division, the first level administrative unit of Bangladesh, and accordingly we found seroprevalence estimates by division.

Seroprevalence estimates in the national survey population accounted for the cluster sample design by adjusting the standard error by the finite population correction factor\textsuperscript{18,19}. Confidence intervals for seroprevalence estimates were calculated with the logit method\textsuperscript{20}. We performed a logit transformation of the proportion to produce the confidence interval, then computed a Wald-type interval on the log-odds scale, which was then transformed to the probability scale. This method is similar to when correcting variance for small sample sizes.

To test whether subgroup seroprevalence estimates for the hospital survey and the national survey were homogenous we performed a Pearson’s chi-squared test adjusted by a method proposed by Rao and Scott\textsuperscript{21}. Rao and Scott’s first order correction takes into account design effects of the survey design used in the nationally representative survey. First order design correction depends only on design effects of the proportion estimates and the marginal proportion estimates.

For the division estimates, we also adjusted the hospital population by age and gender. We were concerned about over- and under- representation of some subgroups in the hospital population, particularly after being stratified into divisions. We applied adjustment weights to correct for this lack of representativeness. We used an iterative proportional fitting procedure, or raking, to find the appropriate weights\textsuperscript{22}. Afterwards, the adjusted division seroprevalence estimates for the hospital survey was compared with the national survey division estimates.

To assess the comparability of inferences about risk factors that would be drawn from each population, we compared the odds of seropositivity in males, controlling for age and division. These comparisons can give us a sense of the overall risk profile of the two populations.

*Sensitivity Analysis*
To assess the sensitivity of the hospital seroprevalence estimates by year of observation, we separated the population into two groups based on hospital admission year. One group was comprised of individuals admitted between 2014 and 2015, while the other was between 2016 and 2017. We then re-estimated seroprevalence estimates stratified by age, gender, and division and compared them.

**Human Subjects Research**

All participants provided informed written consent prior to participation in both studies. In the hospital surveillance study, physicians sought written consent from the patients or their guardians to enroll them in the study. This study's protocol was reviewed and approved by the institutional review board of the icddr,b (Protocol # PR-14060).

In the nationally representative serosurvey, all adult participants provided written, informed consent after receiving detailed explanation of the study. A parent or guardian of all child participants provided written, informed consent on their behalf. The U.S. Centers for Disease Control and Prevention relied on icddr,b's ethical review board approval.

**Results**

In total, 3,536 individuals were included in this analysis. The nationally representative survey started with 2,288 individuals over the age of 14; of them, 16 individuals were excluded for missing data. The hospital survey started with 1,925 individuals aged 14 and over. Of those individuals, 661 were positive for anti-HEV IgM and therefore excluded from the analysis.

Seroprevalence estimates stratified by sex from the hospital data and the national data were notably similar (Table 1). In the national survey, males had a higher seroprevalence than females at 31% seropositivity (95% CI: 26% - 35%) compared to 20% (95% CI: 15% - 25%). Similarly, in the hospital survey males had a seroprevalence of 35% (95% CI: 32% - 38%) and females 21% (95% CI: 18% - 25%). Seroprevalence estimates by sex were consistent between the two data sources, with adjusted chi-square values that indicate no significant difference for both the male and female estimates. Estimated odds ratio (ORs) for the association between being male and having IgG antibodies were
consistent between the two sample groups: hospital OR was 1.8 (95% CI: 1.35-2.42) and the national survey OR was 1.75 (95% CI 1.44-2.12). (Table 2).

In nearly every age group, there were no statistically significant differences between the national and hospital surveys. All age groups except for the 30-39 age group had very low adjusted \( \chi^2 \) statistics and p-values>0.05, indicating that there were no significant differences in the estimates (Table 1). The one age group that did not share this finding was the 30-39 age group, which had a \( \chi^2 \) statistics of 5.11 and a p-value<0.05, indicating that the seroprevalence estimates from the two data sources were significantly different.

The patterns of age-specific seroprevalence were similar in both surveys. In the hospital survey, seroprevalence estimates increased by age, starting at 12% (95% CI: 8%-16%) in the 14-19 age group, reaching 40% (95% CI: 33%-47%) in the 30-39 age group, and finally at 47% (95% CI: 33%-61%) in the 70+ age group. In the national survey, seroprevalence estimates were lower on a whole, but not significantly so. Estimates start at 12% (95% CI: 7%-16%) in the 14-19 age group, peak at 37% (95% CI: 27%-46%) in the 60-69 age group, and finally end at 36% (95% CI: 26%-46%) in the 70+ age group. Figure 2 graphs the seroprevalence estimates by age, and one can see the overlap of confidence intervals in all groups but the 30-39 age group.

In Figure 2a, we have plotted the seroprevalence age distribution stratified by gender for each data source. Initial inspection suggests similarity between the data sources, which is further informed by Figures 2b and 2c which plot male and female groups separately. Both figures 2b and 2c show age subgroup seroprevalence estimates from the hospital survey and the national survey have overlapping confidence intervals.

In four out of the five divisions compared, there were no statistically significant differences in seroprevalence estimates between the hospital survey and national survey (Table 1). All divisions except for the Barisal division had low \( \chi^2 \) statistics and p-values>0.05, indicating that there were no significant differences in the estimates. In the Barisal division, the national survey estimate was 12.7% greater than the hospital survey \( (\chi^2 = 4.39, pvalue = 0.03) \).

We found that the sex distributions of the national survey population and the hospital survey population diverged, with men greatly over-represented in the hospital population at 66% and women only 34% of the population (Table 3). On the other hand, in
the national survey males (47%) and females (53%) were relatively equally represented. Though the overall age distributions of the two datasets were fairly similar to each other, with every age group’s share in the hospital and national population within 4% of each other, age distribution stratified by division was not. Using a visual approximation in Figure 3, the age distribution of the hospital survey clearly varied across divisions and with their national survey counterparts.

Adjusting the hospital survey population by age and sex resulted in lower estimates for all divisions, though not significantly so. (Figure 3a). When comparing the adjusted hospital seroprevalence estimates with those of the national survey, we did not see a significant change in the relationships between the two data sources—there continued to be no significant differences in the seroprevalence estimates between the hospital and population survey for four out of the five division. The statistically significant difference in the Barisal division actually increased in magnitude, with the national survey estimate 15.6% greater than the hospital estimate, compared to the unadjusted difference of 12.7% (Figure 3c).

In the sensitivity analysis, hospital seroprevalence estimates were consistent between time periods in most subgroups (Table 4). In comparing the two time periods 2014-2015 and 2016-2017, estimates for nearly every age group were close and fell within the 95% confidence intervals of each other. The one exception was the 40-49 age group, where seroprevalence estimates were significantly lower for the time period 2014-2015 compared to 2016-2017. Comparing each two-year time period to the overall study found that all age and division estimates for the two time periods fell within the 95% confidence interval of the overall dataset’s estimates.

Discussion

HEV seroprevalence estimates from the hospital serosurvey and the national population survey were largely similar despite the clear distinctions in sampling methodology for each survey. When stratified by age and gender, seroprevalence subgroup estimates had no statistically significant differences, demonstrating that the estimates derived from hospital data looked to be very representative of those derived from the national survey.
Some subgroups did have significant differences between the hospital survey and the national survey. In the 30-39 year old age group, the hospital survey's estimate was 10% greater than that of the national survey ($\chi^2 = 5.11, p \text{ value} = 0.03$); however, this difference did not bear out in the sex specific estimates of the same age group. In the seroprevalence estimates for the Barisal division, the national survey estimate was significantly greater than that of the hospital survey even after adjusting by age and sex. Though the other four out of five divisions had no significant difference, the rank order of division seroprevalence estimates was not consistent between the two data sources.

Men had a similar degree of increased odds of seropositivity in both populations, which is consistent with previous HEV studies finding male preponderance in countries of high endemicity\textsuperscript{5,23,24}. Taken as a whole, this evidence suggests that seroprevalence estimates and odds ratios derived from a hospital-based serosurvey may be good proximate measures for overall population seroprevalence and risk factors.

The sensitivity analysis of the hospital data found that the subgroup seroprevalence estimates for the two time periods were consistent with each other and with the overall dataset. Neither time periods had regularly lower or higher seroprevalence estimates than the other, indicating that the hospital estimates were robust from year to year. Had we only had hospital surveillance data for two years, we would have come to similar conclusions as what we found with the entire dataset. This could mean samples from one or two years of hospital surveillance are sufficient for making population inferences about past infection.

The findings of this research could have useful and far reaching applications to other settings and diseases. Countries with high HEV endemicity but few resources to regularly carry out nationally representative serosurveys could use existing hospital surveillance infrastructure to monitor HEV. Hospitals and clinics that are already designated as sentinel surveillance facilities can be approached to implement research protocols for biobanking samples from eligible patients. With physicians and healthcare workers already familiar with sample collection and storage, deploying a hospital-based serosurvey can be very resource efficient.

For example India does not have a robust case-based surveillance system for viral HEV in place despite having a very high burden of HEV\textsuperscript{2,23,25}. Current surveillance is of aggregated data that does not include information about age distribution or other
demographic characteristics. A hospital-based serosurvey like the one studied in Bangladesh could provide insight into age-specific disease dynamics and assess outbreak risk based on population level immunity.

Communities of displaced persons are another prime candidate for this surveillance strategy. Refugee and displaced persons camps in South Sudan and in Kenya have had protracted HEV outbreaks in recent years, possibly due to overcrowding and poor water conditions. Serological surveys are used to assess relative risk of exposure in the community and may inform approaches to effective HEV vaccination that could be used to prevent or reduce an outbreak. A hospital or clinic-based approach could provide regular HEV surveillance without requiring door-to-door surveying of the community.

This method of using hospital-based convenience samples to infer general population immunity can be especially useful for new and emerging infections such as the novel SARS-CoV-2 virus that do not yet have established surveillance systems. Antibody seroprevalence studies are urgently needed to understand how widespread asymptomatic infections are and what level of immunity is present in the population. Carefully planned representative serosurveys may not be deployable within a few weeks, particularly while social distancing measures are in place. In such cases, convenience samples can be one of few options for researchers to conduct crucial research. For example in a recent paper of Santa Clara County currently in preprint, the study population was recruited from targeted online Facebook ads and required participants to drive to a test site to collect serum samples. Though timely, this sampling method brings up issues of self-selection bias, where participants who have had suspected exposures are more likely to respond to the Facebook recruitment ad. A different strategy with a convenience sample could be to use biobanked samples of hospital patients who were admitted for reasons unrelated to COVID-19. This approach may reduce the likelihood of self-selection bias and diversify the sampling population to include those not using social media or without access to a personal vehicle.

Our understanding of infections that are widespread but under-reported could benefit greatly from hospital-based convenience sampling. Like HEV, leptospirosis is a water borne infection that is highly endemic in Bangladesh and without a regular surveillance system in place, despite the risk of severe renal disease and mortality. A similar hospital-
based serosurvey for leptospirosis, or even one using the same biobanked samples from this HEV study, could be a resource efficient method to further understanding this under-surveilled infection. Leptospirosis is a zoonotic infection spread by the urine of infected animal vectors (most often rodents and canines) that subsequently contaminate water in contact with humans. Understanding regional leptospirosis exposure and susceptibility may greatly inform focused prevention strategies, particularly in the form of vector control and animal vaccination\textsuperscript{40,41}.

There are a number of limitations in this study. Analyses of some subgroups, particularly within the female population, in both the hospital and the national data were limited in statistical power by small sample size. In the case of division subgroups, we did not have enough data to make comparisons for two of the seven divisions. We were not able to assess the full age-specific distribution of HEV seroprevalence because the hospital data did not enroll any patients under the age of 14. The lack of data on the youngest section of the population prevents us from seeing the complete picture of HEV exposure and limits our comparison. Without knowing the seroprevalence in children, it is much more difficult to identify the recent trends in force of infection that are most relevant to stakeholders.

This is particularly pertinent for HEV, because the age dynamics of HEV diverge from similar infections like hepatitis A (HAV)\textsuperscript{23,42}. Unlike HAV seroprevalence which peaks by late childhood, HEV seroprevalence is relatively low in children under the age of 15, and peaks in early to mid-adulthood aged 20 to 35\textsuperscript{43}. In most studies, a low seroprevalence in children is a characteristic of HEV but we were not able to assess this in the hospital-based population\textsuperscript{26,44}.

Finally, this study relied on the detection of anti-HEV IgM antibodies to identify hospitalized individuals with active HEV infection so that they could be excluded from the study. Misclassification during this stage could result in individuals with recent infection being included in the study, which may yield overestimated seroprevalence in younger age groups. Active HEV infection is more prevalent in older children to younger adults (16-35), which can be inferred because the greatest increases in seroprevalence occurs during those years\textsuperscript{23,42}. A model of global burden of HEV also estimated that the average age of infection was 21.1 years old in Asia, so we may expect high levels of anti-HEV IgM around that age\textsuperscript{2}. 

11
If there were numerous instances of this kind of misclassification, we may have found our seroprevalence estimates in the younger age groups to be unduly inflated. Though this misclassification is possible, in our results there was no indication of any irregularities, as younger age groups did not have significantly higher estimates in the hospital survey population.

This study provides a simple comparison between HEV seroprevalence inferences derived from a hospital-based convenience sample and a nationally representative sample. Data from this study suggest that a hospital-based convenience sample accurately reflected the disease dynamics of population-based HEV seroprevalence occurring in the general population. Our findings suggest that this strategy should be considered to understand—at least in general outlines—population-level disease susceptibility for other diseases and in other countries where a nationally representative survey is not possible.
Figure 1. Number of HEV IgM negative patients enrolled in the hospital study aggregated by district and spatial points of the six surveillance hospital locations.
Table 1. Anti-HEV IgG positive seroprevalence by gender, age group, and division subgroups derived from hospital surveillance and national population survey in Bangladesh 2014-2017

<table>
<thead>
<tr>
<th>National Survey</th>
<th>Hospital Survey</th>
<th>$\chi^2$, p-value</th>
<th>Difference in Seroprevalence, National-Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seropositive /Total</td>
<td>Seroprevalence (95% CI)</td>
<td>Seropositive /Total</td>
<td>Seroprevalence (95% CI)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-19</td>
<td>54/453</td>
<td>11.9 (7-16)</td>
<td>25/217</td>
</tr>
<tr>
<td>20-29</td>
<td>94/496</td>
<td>19 (14-23)</td>
<td>69/333</td>
</tr>
<tr>
<td>30-39</td>
<td>127/440</td>
<td>28.8 (22-35)</td>
<td>82/207</td>
</tr>
<tr>
<td>40-49</td>
<td>123/371</td>
<td>33.2 (26-40)</td>
<td>69/171</td>
</tr>
<tr>
<td>50-59</td>
<td>88/260</td>
<td>33.8 (26-41)</td>
<td>63/156</td>
</tr>
<tr>
<td>60-69</td>
<td>61/165</td>
<td>37 (27-46)</td>
<td>52/129</td>
</tr>
<tr>
<td>70+</td>
<td>31/87</td>
<td>35.6 (26-46)</td>
<td>24/51</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>247/1209</td>
<td>20.4 (16--25)</td>
<td>93/436</td>
</tr>
<tr>
<td>Male</td>
<td>331/1063</td>
<td>31.1 (26--35)</td>
<td>291/828</td>
</tr>
<tr>
<td>Division</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barisal</td>
<td>64/136</td>
<td>47.1 (37-57)</td>
<td>67/195</td>
</tr>
<tr>
<td>Chittagong</td>
<td>143/603</td>
<td>23.7 (18-29)</td>
<td>82/292</td>
</tr>
<tr>
<td>Dhaka</td>
<td>189/565</td>
<td>33.5 (21-45)</td>
<td>116/253</td>
</tr>
<tr>
<td>Khulna</td>
<td>40/263</td>
<td>15.2 (8-22)</td>
<td>1/3</td>
</tr>
<tr>
<td>Rajshahi</td>
<td>62/253</td>
<td>24.5 (13-35)</td>
<td>52/306</td>
</tr>
<tr>
<td>Rangpur</td>
<td>45/349</td>
<td>13 (7-18)</td>
<td>5/21</td>
</tr>
<tr>
<td>Sylhet</td>
<td>35/103</td>
<td>34.1 (26-41)</td>
<td>61/194</td>
</tr>
</tbody>
</table>

* Pearson’s $\chi^2$ with Rao and Scott adjustment
* p-value<0.05
Figure 2. Anti-HEV IgG seropositive prevalence by age group derived from hospital survey and national survey of Bangladesh

Figure 2a. Anti-HEV IgG seropositive seroprevalence stratified by gender derived from hospital survey and national survey of Bangladesh
Figure 2b. Anti-HEV IgG seropositive seroprevalence stratified by age group in males derived from both sources

![Graph showing seroprevalence for males by age with data points and error bars.]

Figure 2c. Anti-HEV IgG seropositive seroprevalence stratified by age group in females derived from both sources

![Graph showing seroprevalence for females by age with data points and error bars.]

Table 2. Comparison of odds ratios for males (indicated by yellow dot) controlling for age and division

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Hospital Survey</th>
<th>National Survey</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds Ratios</td>
<td>CI</td>
<td>p</td>
<td>Odds Ratios</td>
<td>CI</td>
<td>p</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.12</td>
<td>0.08 – 0.19</td>
<td>&lt;0.001</td>
<td>0.26</td>
<td>0.15 – 0.47</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male</td>
<td>1.80</td>
<td>1.35 – 2.42</td>
<td>&lt;0.001</td>
<td>1.75</td>
<td>1.44 – 2.12</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>1.03</td>
<td>1.03 – 1.04</td>
<td>&lt;0.001</td>
<td>1.03</td>
<td>1.02 – 1.03</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Chittagong Division</td>
<td>0.61</td>
<td>0.40 – 0.93</td>
<td>0.023</td>
<td>0.32</td>
<td>0.18 – 0.57</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Dhaka Division</td>
<td>1.34</td>
<td>0.89 – 2.04</td>
<td>0.162</td>
<td>0.56</td>
<td>0.26 – 1.19</td>
<td>0.135</td>
</tr>
<tr>
<td>Khulna Division</td>
<td>0.76</td>
<td>0.03 – 8.79</td>
<td>0.827</td>
<td>0.18</td>
<td>0.09 – 0.37</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rajshahi Division</td>
<td>0.31</td>
<td>0.20 – 0.49</td>
<td>&lt;0.001</td>
<td>0.34</td>
<td>0.15 – 0.74</td>
<td>0.009</td>
</tr>
<tr>
<td>Rangpur Division</td>
<td>0.58</td>
<td>0.18 – 1.63</td>
<td>0.332</td>
<td>0.15</td>
<td>0.08 – 0.28</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Sylhet Division</td>
<td>0.60</td>
<td>0.38 – 0.94</td>
<td>0.028</td>
<td>0.57</td>
<td>0.29 – 1.11</td>
<td>0.104</td>
</tr>
<tr>
<td>Observations</td>
<td>1264</td>
<td>2272</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Age and Sex Distribution in Each Data Source

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Hospital %</th>
<th>National %</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-19</td>
<td>17.5</td>
<td>21.1</td>
</tr>
<tr>
<td>20-29</td>
<td>26.0</td>
<td>22.0</td>
</tr>
<tr>
<td>30-39</td>
<td>16.1</td>
<td>18.7</td>
</tr>
<tr>
<td>40-49</td>
<td>13.5</td>
<td>15.8</td>
</tr>
<tr>
<td>50-59</td>
<td>12.5</td>
<td>10.9</td>
</tr>
<tr>
<td>60-69</td>
<td>10.3</td>
<td>7.4</td>
</tr>
<tr>
<td>70+</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Male</td>
<td>65.8</td>
<td>47.1</td>
</tr>
<tr>
<td>Female</td>
<td>34.1</td>
<td>52.9</td>
</tr>
</tbody>
</table>

Figure 3. Comparison of the age distribution of population by division in the national survey and hospital survey
Figure 3a. Hospital seroprevalence by division, comparing sex and age adjusted and unadjusted estimates

Figure 3b. Comparing unadjusted seroprevalence estimates from hospital survey to seroprevalence estimates from national survey, stratified by division

Figure 3c. Comparing age and sex adjusted seroprevalence estimates from hospital survey to seroprevalence estimates from national survey, stratified by division
Table 4. Seroprevalence estimate comparison between complete hospital population, and hospital populations admitted between 2014 and 2015 and between 2016 and 2017

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Seropositive /Total</td>
<td>Seroprevalence (95% CI)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-19</td>
<td>25/217</td>
<td>12 (8-16)</td>
</tr>
<tr>
<td>30-39</td>
<td>82/207</td>
<td>40 (33-47)</td>
</tr>
<tr>
<td>50-59</td>
<td>63/156</td>
<td>40 (33-49)</td>
</tr>
<tr>
<td>60-69</td>
<td>52/129</td>
<td>40 (32-49)</td>
</tr>
<tr>
<td>70+</td>
<td>24/51</td>
<td>47 (33-61)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>93/436</td>
<td>21 (18-25)</td>
</tr>
<tr>
<td>Male</td>
<td>291/828</td>
<td>35 (32-38)</td>
</tr>
<tr>
<td><strong>Division</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barisal</td>
<td>67/195</td>
<td>34 (28-41)</td>
</tr>
<tr>
<td>Chittagong</td>
<td>82/292</td>
<td>28 (23-34)</td>
</tr>
<tr>
<td>Dhaka</td>
<td>116/253</td>
<td>46 (40-52)</td>
</tr>
<tr>
<td>Rajshahi</td>
<td>52/306</td>
<td>17 (13-22)</td>
</tr>
<tr>
<td>Sylhet</td>
<td>61/194</td>
<td>31 (25-38)</td>
</tr>
</tbody>
</table>
Supplementary Figure 1. Hospital seroprevalence by sex and admission year, 2014 to 2015, 2016 to 2017, and the entire study period

Supplementary Figure 2. Hospital seroprevalence by age group and admission year, 2014 to 2015, 2016 to 2017, and the entire study period
Supplementary Figure 3. Hospital seroprevalence by division and admission Year, 2014 to 2015, 2016 to 2017, and the entire study period
References


EDUCATION

**Johns Hopkins Bloomberg School of Public Health**
Master of Science (ScM) in Epidemiology
Concentration: Infectious Disease, Expected May 2020. GPA: 3.6
Baltimore, MD

**University of California, Berkeley**
Bachelor of Arts in Public Health
Conferred May 2017. GPA: 3.8 with Honors Research Thesis
Berkeley, CA

WORK EXPERIENCE

**Graduate Researcher at Bloomberg School of Public Health**
International Vaccine Access Center
September 2018—April 2019
Baltimore, MD

- Identified risk factors for low immunization coverage to help inform ongoing national vaccination campaign by Indian health ministry
- Mapped spatial clustering of subpopulations with low immunization coverage at high risk for disease outbreak
- Authored subset of policy recommendations on immunization risk factors for state-level health departments

**Research Data Analyst at UC Berkeley School of Public Health**
Remais Infectious Disease and Environmental Change Research Group
May 2017—June 2018
Berkeley, CA

- Coordinated research study activities both within the research group and with outside collaborators
- Authored policy recommendations for partners at the CA Department of Health on the state’s future projected risk of West Nile Virus and Valley Fever given climate change projections of temperature and precipitation
- Assisted launch of investigation on industrial waste and flood risk in California by compiling data and presenting exploratory spatial analysis findings

**Environmental Health Intern at the CDC**
June—August 2016
Atlanta, GA

- Delivered critical data analysis support for a water contamination investigation of Superfund sites
- 1 of 11 selected interns for environmental health leadership training at the CDC

**Emerging Leaders Internship at GLIDE Memorial**
HIV/Hep C and Harm Reduction Services
June—August 2014
San Francisco, CA

- Provided HIV/Hep C testing, street outreach, and overdose prevention training within high risk populations
- Facilitated weekly critical discussions within internship program on community health and social justice issues
**RESEARCH EXPERIENCE**

<table>
<thead>
<tr>
<th><strong>Graduate Researcher</strong></th>
<th>April 2019—Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bloomberg School of Public Health</td>
<td></td>
</tr>
<tr>
<td>• Comparing hepatitis E seroprevalence estimates in Bangladesh derived from a nationally representative sero-survey with estimates from hospital surveillance system</td>
<td></td>
</tr>
<tr>
<td>• Working with large national datasets of population serology to estimate prevalence of hepatitis E in Bangladesh</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Undergraduate Research Apprentice</strong></th>
<th>September 2016—May 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>UC Berkeley School of Public Health</td>
<td></td>
</tr>
<tr>
<td>• Modeled relationship between leptospirosis, a water borne disease, and land-cover change in China</td>
<td></td>
</tr>
<tr>
<td>• Developed high resolution maps of land cover change to illustrate deforestation and urbanization trends</td>
<td></td>
</tr>
</tbody>
</table>

**SKILLS & MISC**

*Computing and Programming:* Highly proficient with R, STATA, ArcGIS, Excel, REDCap, SQL

*Teaching:* Former ceramics instructor, TA for undergraduate Intro to Public Health course

*Languages:* Fluent Korean, Conversational Mandarin