

1 **Annual risk of hepatitis E virus infection and seroreversion: insights from**
2 **a serological cohort in Sitakunda, Bangladesh**

3 Amy Dighe*¹, Ashraful Islam Khan², Taufiqur Rahman Bhuiyan², Md Taufiqul Islam², Zahid Hasan Khan², Ishtiakul
4 Islam Khan², Juan Dent Hulse¹, Shakeel Ahmed³, Mamunur Rashid³, Md Zakir Hossain³, Rumana Rashid³, Sonia
5 Hegde¹, Emily S Gurley^{†1}, Firdausi Qadri^{†2}, Andrew S Azman^{†1,4,5}

6 1. Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.

7 2. icddr,b, Dhaka, Bangladesh.

8 3. Bangladesh Institute of Tropical and Infectious Diseases, Chattogram, Bangladesh

9 4. Geneva Centre for Emerging Viral Diseases, Geneva University Hospitals, Geneva, Switzerland

10 5. Division of Tropical and Humanitarian Medicine, Geneva University Hospitals, Geneva, Switzerland

11 *Corresponding author: adighe1@jhmi.edu, Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe St,

12 Baltimore, MD 21205

13 †Equal contributions

14

15

16

17

18

19

20

21 **Abstract**

22 Hepatitis E virus (HEV) is a major cause of acute jaundice in South Asia. Gaps in our understanding of transmission
23 are driven by non-specific symptoms and scarcity of diagnostics, impeding rational control strategies. In this context,
24 serological data can provide important proxy measures of infection. We enrolled a population-representative
25 serological cohort of 2337 individuals in Sitakunda, Bangladesh. We estimated the annual risks of HEV infection and
26 seroreversion both using serostatus changes between paired serum samples collected 9 months apart, and by fitting
27 catalytic models to the age-stratified cross-sectional seroprevalence. At baseline, 15% (95CI: 14-17%) of people
28 were seropositive, with seroprevalence highest in the relatively urban south. During the study, 27 individuals
29 seroreverted (annual seroreversion risk: 15%, 95CI: 10-21%), and 38 seroconverted (annual infection risk: 3%, 95CI:
30 2-5%). Relying on cross-sectional seroprevalence data alone, and ignoring seroreversion, underestimated the annual
31 infection risk fivefold (0.6%, 95CrI: 0.5-0.6%). When we accounted for the observed seroreversion in a reversible
32 catalytic model, infection risk was more consistent with measured seroincidence. Our results quantify HEV infection
33 risk in Sitakunda and highlight the importance of accounting for seroreversion when estimating infection incidence
34 from cross-sectional seroprevalence data.

35

36

37

38

39

40

41

42 Introduction

43 Hepatitis E virus (HEV) is an important cause of acute jaundice in populations with limited access to safe drinking
44 water across South Asia and Africa [1–5]. Transmitted from person-to-person via faecal-contaminated water, HEV
45 genotypes 1 and 2 (HEV-1 and 2) carry a high risk of death if contracted during pregnancy [6] and regularly cause
46 outbreaks, with particularly long-lasting transmission documented in displaced persons camps [7]. In Bangladesh,
47 HEV is the leading cause of hospital-attended acute jaundice [3]. Although more recent data are lacking, an analysis
48 of verbal-autopsies from 1998-2007 estimated that 19-25% of maternal deaths were associated with jaundice,
49 suggesting that HEV may be a key cause of maternal mortality nationally [2].

50 The risk factors driving HEV infection are not fully understood and attempts to slow transmission through emergency
51 water and sanitation interventions have had limited success [8]. In the absence of effective treatment, vaccination is a
52 promising tool to avert cases and deaths. Whilst an efficacious vaccine exists [9], the lack of reliable burden
53 estimates is one of several barriers preventing the World Health Organisation (WHO) from recommending its routine
54 use [10,11]. Sparse surveillance data and differing model assumptions mean that estimates of morbidity and mortality
55 attributable to HEV vary widely, rendering them difficult to interpret [4,5,12,13]. Understanding the drivers of HEV
56 infection and improving incidence estimates can greatly support both global and local decision makers.

57 Although not a direct measure of disease incidence, serological data can provide important proxy measures of
58 infection. Age-stratified cross-sectional seroprevalence is often used to estimate the rate at which seronegative
59 individuals become infected with a pathogen [14] and has previously been used to estimate HEV infection incidence
60 though past approaches have had several limitations [4,5]. Despite evidence that antibodies to HEV wane over time
61 [15,16], models have ignored seroreversion, which cannot always be reliably estimated from cross-sectional
62 seroprevalence [17]. Additionally, such approaches have assumed infection risk does not vary with age or time,
63 which may not hold for HEV. Collecting serum samples from the same individuals at different time-points can
64 overcome some of these limitations with observed serostatus changes providing a direct measure of seroincidence
65 and seroreversion.

66 In this study, we aimed to fill several gaps in our understanding of HEV infections by enrolling a population-
67 representative longitudinal serological cohort of 580 households in an HEV-1 endemic region in Bangladesh [18,19]
68 to explore infection risk factors and estimate the annual risk of infection, and the rate of seroreversion. As a
69 secondary goal we aimed to compare the concordance of estimates of infection incidence derived from cross-
70 sectional data to those observed in the longitudinal cohort.

71 **Methods**

72 Serosurvey design

73 We tested serum samples from a population-representative cohort recruited between March 2021 and February 2022
74 with the original aim of estimating *Vibrio cholerae* O1 seroincidence in the Sitakunda sub-district of Chattogram,
75 Bangladesh (Figure S1). Households were recruited through a previously described two-stage sampling process [20].
76 Briefly, we first divided Sitakunda into 1km² grid cells and randomly sampled cells weighted by the number of
77 household structures identified by satellite imagery. We then randomly sampled structures to visit within each grid
78 cell weighted by whether they were single- or multi-story units. For each household, we sought consent from the
79 household head and attempted to enrol all members ≥ 1 years old. Study staff administered a questionnaire covering
80 household-level infrastructure, assets and sanitation facilities to household representatives, and an individual-level
81 questionnaire on demographics and drinking-water sources to all consenting household members. In addition, ~5ml
82 of venous blood was collected from each consenting household member (~3ml for those <5yrs). Enrolled households
83 were visited at a subsequent timepoint approximately 9 (range 7-11) months from baseline, to ask follow-up
84 questions and repeat blood collection.

85 Sample testing

86 Paired serum samples from the two survey rounds were tested for anti-HEV immunoglobulin (IgG) at icddr,b using
87 commercially available Wantai HEV IgG ELISA kits (Wantai Biological, China). Following the manufacturer's
88 instructions, samples with a standardized optical density >1.1 were considered seropositive, those <0.9 were
89 considered negative and those between 0.9-1.1 borderline. Borderline results were excluded from analyses.

90 Statistical analysis

91 *Seroprevalence and risk factor analyses*

92 We estimated baseline seroprevalence and 95% confidence intervals, accounting for household sampling survey
93 design, using the Rao-Scott method, implemented in the R package “survey” [21]. We created smoothed maps of
94 household seroprevalence using inverse distance weighting and assessed the spatial autocorrelation by estimating
95 the semivariogram for household seroprevalence. We explored the relationship between individual- and household-
96 level variables and baseline seropositivity using mixed effect logistic regression models to account for household-
97 level random effects. Firstly, we estimated univariate odds ratios of seropositivity for variables pertaining to
98 demography, drinking-water and sanitation, history of jaundice and livestock keeping. We then constructed a
99 multivariable model including those variables that were statistically significant ($p \leq 0.05$) in univariate analyses and
100 those identified *a priori* as potentially causally related to HEV exposures.

101 Drinking-water sources and sanitation facilities were categorized as improved and unimproved based on definitions
102 from the WHO/UNICEF Joint Monitoring Programme (JMP) service ladders for drinking water [22] and sanitation
103 facilities [23].

104 *Annual risks of infection and seroreversion*

105 We estimated the annual risk of infection of seronegative individuals (from here on referred to as just the annual risk
106 of infection) and seroreversion of seropositive individuals using two methods: (1) using the observed changes in
107 seroprevalence between study visits, and (2) fitting catalytic models to age-stratified seroprevalence data from a
108 single study visit.

109 The annual risk of infection was first estimated by dividing the number of individuals who seroconverted over the
110 course of the study period, n_{sc} , (i.e., baseline seronegative individuals who became seropositive) by the total number
111 of person-time at risk during the study period (total person-time for those who remained seronegative throughout,
112 pt_{sn} , plus half the person-time for those who seroconverted, pt_{sc}) as described by equation (1). Similarly, the annual
113 risk of seroreversion was first estimated by dividing the number of baseline seropositive individuals who became

114 seronegative, n_{sr} , by the total person-time at risk (total person-time for those who remained seropositive throughout
115 pt_{sp} , plus half the person-time for those who seroreverted, pt_{sr}) as described by equation (2). We assumed on
116 average individuals seroconverted or seroreverted at the midpoint of their time in the study.

117
$$\text{annual risk of infection} = \frac{n_{sc}}{pt_{sn} + \frac{1}{2}(pt_{sc})} \quad (1)$$

118

119
$$\text{annual risk of seroreversion} = \frac{n_{sr}}{pt_{sp} + \frac{1}{2}(pt_{sr})} \quad (2)$$

120 Alternative estimates of the annual risk of infection (often presented as the rate known as the force of infection when
121 estimated in this way) were produced by fitting two catalytic models of seroconversion to the cross-sectional age-
122 stratified seroprevalence data collected at baseline. In Model 1 we assumed no antibody waning, which has been the
123 traditional approach to estimating HEV infection incidence [4,5]. In Model 2, due to evidence of seroreversion in our
124 empirical data and in previous studies [15,16], we allowed for seroreversion at rate, ρ . Since previous work has
125 shown that the annual risk of infection and the seroreversion rate often cannot be reliably estimated from cross-
126 sectional seroprevalence data simultaneously due to identifiability issues [17], we used our empirical estimates for
127 the rate of seroreversion in Model 2. For both models, we initially considered infection risk to be constant across age
128 and time, then repeated model fitting whilst allowing the risk to vary by age group as has been done previously
129 [24,25]. Model solutions are presented in the Supplementary Material. The seroprevalence data was assumed to be
130 binomially distributed. Models were fitted within a Bayesian framework using a Hamiltonian Monte Carlo algorithm
131 implemented within rstan [26,27]. We used a uniform prior between 0 and 1 for the per capita annual risk of infection,
132 and between 0 and 10 for the per capita annual risk of seroreversion to conservatively include previous measures of
133 HEV antibody persistence which have ranged from months to many years [15,16,28,29]. Model fit was assessed
134 using Leave One Out Cross-Validation (LOO-CV) as implemented in the loo R package [30].

135 To estimate the annual number of HEV infections we multiplied the annual risk of infection, based on observed
136 seroconversions within the cohort, by the estimated population of Sitakunda in 2021. We extrapolated the age-
137 stratified population counts from the 2011 National Census- [31], assuming that the population grew by 1.5% each
138 year between 2011 and 2021. Since we did not include <1-year olds in our survey we subtracted 20% of the
139 population count for 0-4 year olds based on the age distribution presented in the US Census Bureau International
140 Database for Bangladesh in 2021 [32].

141 Ethical review

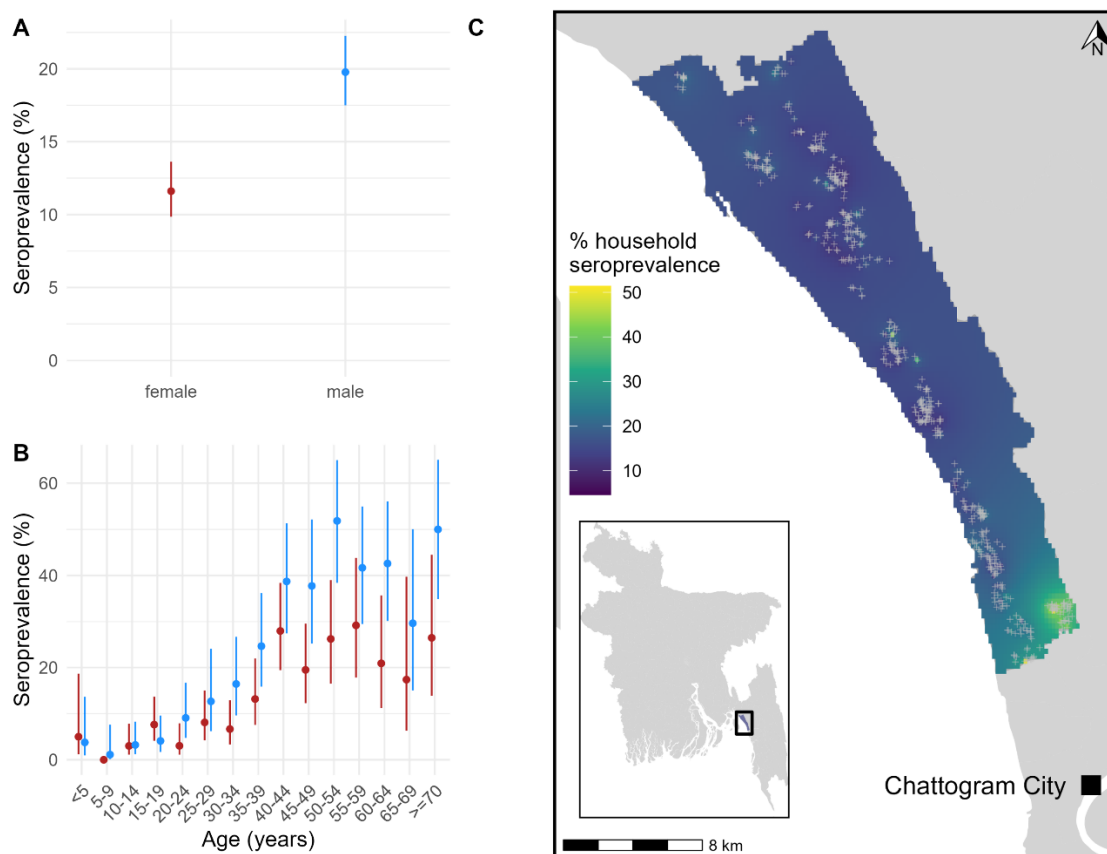
142 Adult study participants provided written, informed consent. Parents or guardians of all participants <18 years were
143 asked to provide consent on their behalf, with those 11-17 years old also providing written assent. The protocols for
144 the original study and extension to test samples for HEV antibodies, were approved by the icddr,b research and
145 ethics review committee and the Johns Hopkins Bloomberg School of Public Health institutional review board.

146 Results

147 *HEV seroprevalence at baseline*

148 Between 27 March and 13 June 2021, 2337 individuals from 580 households were recruited to the serological cohort,
149 and 2301 (98%) were tested for anti-HEV IgG antibodies (Table S1). At enrolment, 15% (95% confidence interval
150 [CI]: 14-17%, ICC: 0.05, design effect: 1.15) of the sampled population of Sitakunda had antibodies indicating past
151 HEV infection (which we refer to as being “seropositive”). Seroprevalence was significantly higher in males (20%,
152 95%CI: 17-22%) than females (12%, 95%CI: 10-14%) (Figure 1A), with the difference becoming more apparent
153 during adulthood (Figure 1B). Seroprevalence was low in children and increased until approximately age 40 before
154 plateauing (Figure 1B). Household seroprevalence was higher in populations living in the relatively urbanized south,
155 near Chattogram city, with one notable cluster in the southeast (Figure 1C). The seroprevalence in this cluster was
156 46% (95%CI: 38-54%), more than 3-fold higher than the average. Compared to just 4% of the overall sample, 20% of
157 individuals within this cluster reported that their primary drinking water source had been unavailable at least once in
158 the month prior to the baseline survey (Table S1). When comparing demographic characteristics, a larger proportion

159 of individuals in the cluster were male (53% compared to 46%), and more households had a monthly income
160 <10,000TK, shared sanitation facilities, and kept livestock (Table S1). While we did detect some areas with elevated
161 seroprevalence, across the Sitakunda, we did not detect strong spatial correlation in household seroprevalence
162 (Figure S2).



163
164 **Figure 1. A.** Baseline seroprevalence by sex, with 95% confidence intervals. **B.** Baseline seroprevalence stratified by
165 age and sex, with 95% confidence intervals. **C.** Smoothed spatial variation in baseline household seroprevalence
166 across Sitakunda. Grey crosses represent the location of sampled households.

167 To further understand the sex differences in seroprevalence and to identify other potential individual- and household-
168 level risk factors for seropositivity, we conducted logistic mixed effects regression analyses. In multivariable models,
169 we found significantly increased odds of seropositivity among those ≥ 40 years, those reporting that their primary
170 water source was unavailable at least once in the past month and those with business or other occupations outside

171 the home (Table 1). Including both sex and occupation in the multivariable model attenuated the effect size for sex
 172 (1.4, 95% CI 1.0-2.0) to the limit of statistical significance.

173 **Table 1. Potential risk factors for past HEV infection at baseline**

	Characteristic	Seropositive	Seronegative	OR (95%CI)	OR (95%CI)
		n (%)	n (%)	Unadjusted	Adjusted
Sex	Female	143 (41%)	1089 (56%)	Reference	
	Male	210 (59%)	852 (44%)	1.9 (1.5, 2.4)	1.4 (1.0, 2.0)
Age in years	<5	4 (1%)	89 (5%)	Reference	
	5-14	9 (3%)	427 (22%)	0.5 (0.1, 1.5)	1.7 (0.3, 10.6)
	15-39	97 (27%)	932 (48%)	2.4 (0.8, 6.7)	5.1 (0.7, 35.6)
	40-100	243 (69%)	493 (25%)	12.9 (4.5, 36.6)	20.8 (2.9, 147.9)
Occupation	Homeworker	127 (36%)	717 (37%)	Reference	
	Business*	138 (39%)	364 (19%)	2.2 (1.7, 3)	1.8 (1.2, 2.7)
	Farmer	27 (8%)	53 (3%)	3.1 (1.8, 5.4)	1.7 (0.9, 3.3)
	Student	20 (6%)	646 (33%)	0.2 (0.1, 0.3)	0.5 (0.3, 1.0)
	Child	5 (1%)	91 (5%)	0.3 (0.1, 0.8)	2.6 (0.4, 15.8)
	None	23 (7%)	48 (2%)	3.0 (1.7, 5.4)	1.3 (0.7, 2.4)
	Other**	13 (4%)	21 (1%)	3.6 (1.7, 7.9)	3.3 (1.4, 7.8)
Travel: time since last leaving the village	> 1 year	322 (91%)	1818 (94%)	Reference	
	1 month - 1 year	22 (6%)	90 (5%)	1.4 (0.8, 2.3)	NA
	1 week - 1 month	5 (1%)	19 (1%)	1.5 (0.5, 4.4)	NA
	< 1 week	4 (1%)	14 (1%)	1.6 (0.5, 5.2)	NA
At least once in the last month primary drinking water source was unavailable	No	328 (93%)	1871 (96%)	Reference	
	Yes	24 (7%)	69 (4%)	2.0 (1.2, 3.4)	2.4 (1.3, 4.4)
Use of piped water as a primary source	No	296 (84%)	1685 (87%)	Reference	
	Yes	57 (16%)	252 (13%)	1.3 (0.9, 1.9)	1.2 (0.8, 1.8)

Use of tubewell as a primary source	No	81 (23%)	385 (20%)	<i>Reference</i>	
	Yes	272 (77%)	1552 (80%)	0.8 (0.6, 1.1)	NA
Use of a public tap/standpipe as a primary source	No	319 (90%)	1790 (92%)	<i>Reference</i>	
	Yes	34 (10%)	147 (8%)	1.3 (0.9, 2.0)	NA
Reported ever having had acute jaundice [†]	No	267 (76%)	1562 (80%)	<i>Reference</i>	
	Yes	10 (3%)	46 (2%)	1.3 (0.6, 2.7)	NA
Household level Type of dwelling	Single house	245 (69%)	1264 (65%)	<i>Reference</i>	
	Several separate structures	40 (11%)	271 (14%)	0.8 (0.5, 1.1)	NA
	Flat in a multi-story building	35 (10%)	175 (9%)	1.0 (0.7, 1.6)	NA
	Flat in a single-story building	22 (6%)	144 (7%)	0.8 (0.5, 1.3)	NA
	Room in a larger dwelling	11 (3%)	87 (4%)	0.7 (0.3, 1.3)	NA
	Household income	<10,000 BDT ^{††}	69 (20%)	298 (15%)	1.4 (1.0, 1.9)
>10,000 BDT ^{††}		284 (80%)	1643 (85%)	<i>Reference</i>	
Sanitation facility category	Improved private	260 (74%)	1476 (76%)	<i>Reference</i>	
	Improved shared	90 (25%)	453 (23%)	1.1 (0.9, 1.5)	NA
	Unimproved	3 (1%)	12 (1%)	1.4 (0.3, 5.6)	NA
Keeps mammalian livestock in the household	No	250 (71%)	1381 (71%)	<i>Reference</i>	
	Yes	103 (29%)	555 (29%)	1.0 (0.8, 1.3)	NA

174 * includes responses “business”, “business man” and business work “outside the home”

175 **when asked to specify, common responses included “service”, “private services”, “teacher” and “driver”.

176 †described as an illness during which the eyes or skin turn yellow. This was only asked at the second time point.

177 ††10,000 BDT is approximately 100 USD

178

179 Self-reported history of jaundice was not associated with significant increased odds of seropositivity. Of the 59 people

180 who reported having ever had acute jaundice (lasting <3 months), nine reported having had jaundice between survey

181 rounds but none seroconverted. Keeping mammalian livestock was not associated with significantly different baseline
182 seropositivity. Only one household kept pigs – a known host of zoonotic HEV genotypes 3 and 4.

183 *Evidence of infection during the study period – empirical estimation of the annual risk of infection*

184 Of the 1580 individuals who were seronegative at baseline and provided blood at the end of the study, 38 became
185 seropositive, with similar rates of seroconversion in men and women. This translates to an annual risk of infection of
186 3% (95% CI: 2-5%) for a seronegative individual in Sitakunda, or approximately 12500 infections in those 1-year-old
187 in 2021. The mean annual risk of infection was higher in adults (4%; 95 CI: 3-6%) than in children <18 years old (3%;
188 95%CI: 1-5%) but this was not statistically significant. Of the 38 seroconverters, seven lived within the high
189 seroprevalence cluster where the annual infection risk was 19% (95%CI: 7-38%). The 38 seroconverters came from
190 29 households, and among these, five households had >1 seroconverter, including one where five members became
191 seropositive.

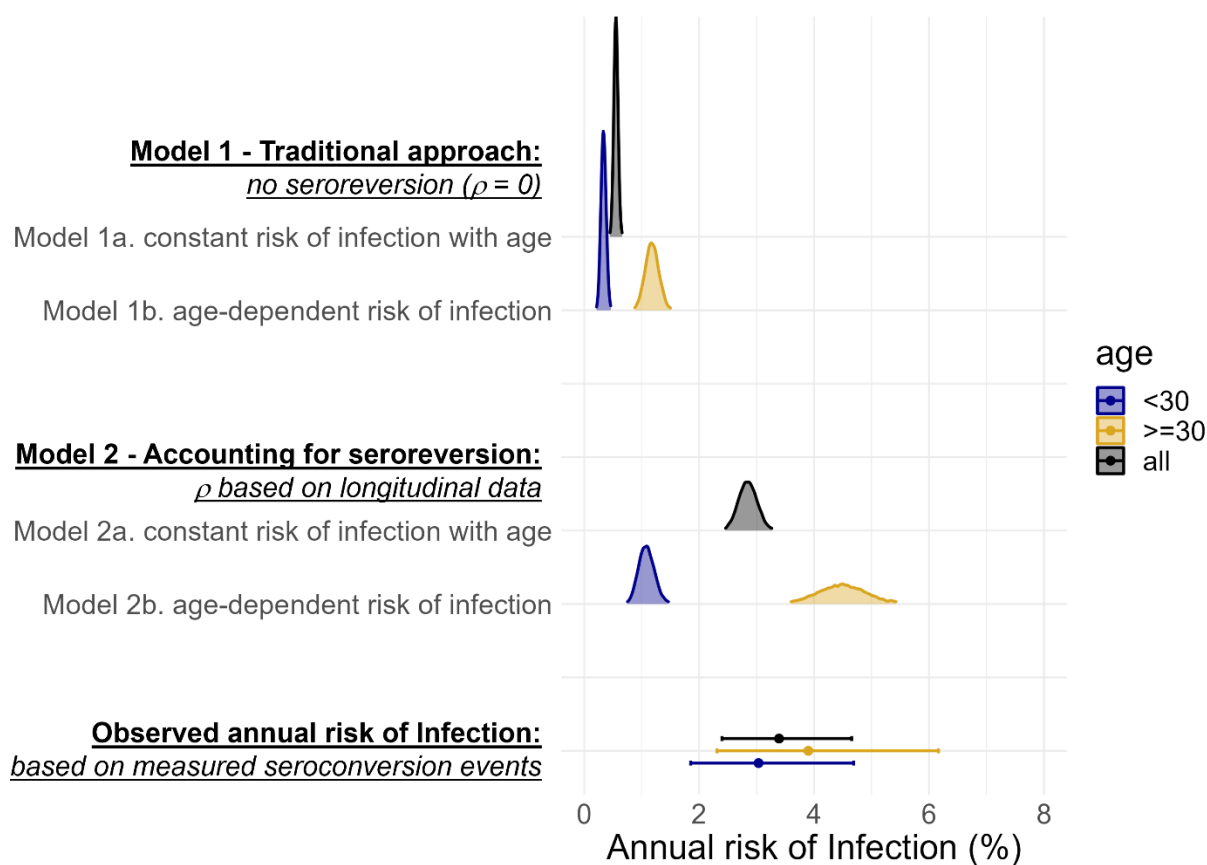
192 *Waning of antibodies during the study period – empirical estimation of the rate of seroreversion*

193 Of the 266 individuals who were seropositive at baseline and for whom we have paired samples, 27 became
194 seronegative during the study. This translates to an annual risk of seroreversion of 15% (95%CI: 10-21%).
195 Seroreversion rates were slightly lower in males (12%, 95%CI: 6-20%) than females (19%, 95%CI: 11-32%).
196 Seroreversion rates were significantly higher in children than in adults (Figure S3). Of the five children <10 who were
197 seropositive at baseline, four seroreverted. The mean time to seroreversion for children <10 years was estimated to
198 be 7 months (95%CI: 3-24 months) compared to 8 years (95%CI: 5-12 years) for those ≥10 years.

199 As changes in seropositivity could result from small fluctuations in antibody concentration or measurement error
200 around the cut-off, we compared the optical density to cut-off ratios for individuals across rounds. The majority of
201 seroreverters and seroconverters had a considerable change in antibody titres (Figure S4A-B). Several individuals
202 classified as seropositive at baseline had a substantial increase in their optical density to cut-off ratio suggesting
203 reinfections (Figure S4.C).

204 *Estimating the annual risk of infection from cross-sectional data*

205 The annual risk of HEV infection has previously been estimated by fitting catalytic models of seroconversion to age-
 206 stratified cross-sectional seroprevalence data, assuming antibodies do not wane. When we fitted a model assuming
 207 lifelong antibody persistence (Model 1) to our baseline age-stratified seroprevalence data, our estimated annual risk
 208 of infection was 5-times lower than what we had measured based on observed seroconversion events (Figure 2).
 209 This same magnitude of difference was seen both when we assumed the risk of infection was the same across ages
 210 (Model 1a), and when we allowed risk of infection to change around 24-30 years of age (Model 1b, which optimized
 211 model fit; Table S3, Figure S5 & S6).



212
 213 **Figure 2.** A comparison of posterior estimates of the annual risk of infection from catalytic models 1 and 2 fitted to
 214 age-stratified cross-sectional data (smoothed density curves) and the estimates of the annual risk of infection from
 215 observed seroconversion events captured in our longitudinal serostatus data (points with bars representing 95% CIs).

216 When we fitted a model (Model 2) assuming antibodies wane below detectable levels at the rate we measured in the
217 cohort (1.8 (95%CI: 0.5-4.6) in <10-year-olds and 0.1 (95%CI: 0.1-0.2) in \geq 10-year-olds), the resulting estimates of
218 annual risk of infection were more comparable to what we measured longitudinally (Figure 2). When we assumed
219 that all age-classes experienced the same risk of infection (Model 2a), the estimated annual risk of infection was
220 2.8% (95%CrI: 2.5-3.2%) similar to the 3.4% (95%CI: 2.4-4.6%) based on the observed seroconversion events.
221 However, this model fit resulted in overestimates of age-specific seroprevalence for younger age-classes and
222 underestimates for older age-classes when compared to the observed data (Figure S5). Allowing age-class-
223 dependent risk of infection (Model 2b) improved model fit, with the best fit achieved when the risk changed at around
224 30-years. For individuals >30, the annual risk of infection was 4.5% (95%CrI: 3.7-5.4%), similar to the measured risk
225 for that age group (3.8%, 95%CI: 2.3-6.1%). The annual risk of infection in those <30-years though was lower than
226 what was measured in the cohort (Figure 2). Using the cross-sectional seroprevalence data from follow-up rather
227 than baseline produced very similar estimates of seroprevalence by age and annual risk of infection (Figure S7 &
228 Table S4).

229 When we attempted to simultaneously fit both the rate of infection and of seroreversion, our estimates were not
230 compatible with the seroreversion rate we measured in the cohort and model fit did not improve (Table S3, Figure
231 S5). When we assumed infection risk was constant with age, the annual risk of seroreversion approached zero
232 (0.09%, 95%CrI: 0.00-0.35%). When we allowed different infection risks for those over and under 30, the annual risk
233 of seroreversion was 5.7% (95%CrI: 2.1-10.7%), less than half what we observed during the study (Figure S8).

234 Discussion

235 In this population-representative longitudinal cohort study, we collected two types of serological data as proxy
236 measures of HEV infection in an endemic region of Bangladesh. We used age-stratified cross-sectional
237 seroprevalence data to explore risk factors associated with past HEV infection and to produce traditional estimates of
238 the annual risk of infection, whilst longitudinal serostatus changes allowed us the opportunity to capture incident
239 infections and seroreversion.

240 At 15% (95%CI: 14-17%), our estimated baseline seroprevalence was similar to the estimate derived from the 2015
241 national serosurvey for Sitakunda (19%, 95%CrI: 12-30%) [33]. We estimated that 3% of seronegative people
242 become infected annually in Sitakunda, translating to approximately 12500 HEV infections in people \geq 1 year old
243 which is approximately half that of previous estimates for rural Bangladesh in 2003-4 (6%, 95%CI: 4-8%) [34]. This
244 could reflect lower transmission intensity in Sitakunda than other rural parts of the country or decreasing incidence
245 over time.

246 Like previous studies in Bangladesh [33,35], we found male sex to be associated with higher odds of anti-HEV
247 seropositivity in univariate analyses. Our results suggest occupation may confound the effect of sex on infection risk,
248 with those working in business or other occupations outside the household significantly more likely to be seropositive
249 and male. Collecting data on water source use outside the household may help us better understand how these
250 consumption patterns are related to HEV infection risk. Despite evidence that HEV outbreaks have been caused by
251 breeches and contamination of municipal piped water supply systems in urban Bangladesh [36], we did not find use
252 of piped water in the week prior to the survey to be associated with significantly higher odds of seropositivity in this
253 study. However, individuals who reported that their primary water source had been unavailable at least once in the
254 last month were more than 2-times more likely to be seropositive, suggesting that use of water lines with insufficient
255 supply, or having to use alternatives, may increase risk of infection.

256 In the absence of longitudinal data, estimating annual HEV infection risk has traditionally relied on fitting catalytic
257 models to cross-sectional seroprevalence, assuming life-long antibody persistence and constant risk of infection. By
258 testing paired samples and capturing seroconversion events, we measured the annual risk of HEV infection in
259 Sitakunda to be 5-times higher than the estimate generated by applying traditional approaches to our baseline
260 seroprevalence data. Using observed seroreversion events to inform a reversible model of seroconversion allowed
261 us to obtain estimates of the annual risk of infection closer to what we measured during our study. Simultaneously
262 fitting the rate of seroreversion and infection risk to cross-sectional seroprevalence data could not eliminate the need
263 for empirical measures of seroreversion, likely due to identifiability issues [17]. We found that allowing for different

264 annual infection risk in those >30 greatly improved the model fit, but it was not possible to determine if this reflects a
265 higher risk in those >30 or lower incidence over the past 30 years.

266 Our estimated annual risk of seroreversion (15%) is higher than existing estimates which were approximately 2% in
267 rural Bangladesh (20% antibody loss over 10-12 years), 4% in Kashmir (50% over 14 years) and 5% in China (30%
268 over 6.5 years) [15,16,29]. We expect this is partly due to the shorter time between paired sample collection in our
269 study compared to previous studies where seroreversion events may have been masked by reinfection. The
270 seroreversion rate was significantly higher in children <10 than in adults – a trend also observed in the previous study
271 of anti-HEV IgG loss in Bangladesh [16]. HEV antibodies were also measured to wane quickly in a study in children
272 in Egypt, becoming undetectable in a matter of months [28]. In our study, most children who seroreverted started with
273 optical density values near the upper limit of the dynamic range of the assay suggesting that children may experience
274 faster antibody waning than adults, rather than mounting a lower initial antibody response. Potential reasons for this
275 observed difference in seroreversion rate are not well understood though could include differences in the immune
276 response to HEV in young children, or less antibody persistence due to fewer repeat infections in children compared
277 to adults who have been at risk for longer. A larger sample of seropositive children is needed to investigate these
278 differences in antibody persistence further and to understand the implications for estimating risk of infection in this
279 age group.

280 Our study has several limitations. Firstly, we used a non-quantitative serological assay and did not use reference
281 serum to allow for generalized comparisons to other studies [37]. We relied on the threshold specified by the kit
282 instructions to classify samples as seropositive and seronegative, assuming perfect assay sensitivity and specificity
283 for detecting infections during the study period. Sensitivity and specificity are estimated to be high [38,39] but the use
284 of this threshold is unlikely to produce perfect, generalisable classification. Although the majority of serostatus
285 changes were associated with large changes in antibody titres, a small minority of seroconversion and reversion
286 events involved relatively small changes. Without a gold standard assay for comparison, we cannot rule out that
287 some serostatus changes could be due to noise. We also saw several large boosts in antibody measures in
288 seropositive individuals that likely represent uncounted reinfections.

289 Secondly, changes in infection risk over time could contribute to the discrepancy between the seroincidence
290 measured longitudinally and our estimates from cross-sectional data. For example, an undetected outbreak could
291 have elevated the risk of infection measured during our study, but very low numbers of people self-reporting jaundice,
292 and similar seroprevalence to other studies suggest this did not occur. Endemic transmission of HEV is not known to
293 be strongly seasonal in Bangladesh [34], but if there is a seasonal component then we could have slightly biased
294 estimates by not spreading equal at-risk person-time across a full year. Spanning 10.5 months, we would not expect
295 the effect to be considerable in our study. Finally, although we were able to estimate the HEV infection rate in
296 Sitakunda, the absence of a serological assay capable of distinguishing between HEV genotypes, with potentially
297 very different clinical consequences, constrains our ability to translate this into estimates of disease burden. To date,
298 genotyped clinical cases in Bangladesh have all been classified as HEV-1 [18,40], but zoonotic HEV likely circulates
299 in pigs in the country [41] and the incidence of spill over is unknown.

300 Our results provide evidence of endemic circulation of HEV in Sitakunda, with people who work in occupations
301 outside the home and those reporting their primary water sources to be recently unavailable at higher risk of
302 infection. In the face of widespread under-reporting of hepatitis E cases, estimates of infection incidence from
303 serological data, while imperfect, are important for improving our understanding of transmission, risk and burden of
304 hepatitis E. We were able to evaluate estimates produced by traditional approaches against empirical estimates of
305 seroincidence, demonstrating the need to account for the rate of antibody waning and differences in the risk of
306 infection experienced by different age groups, to avoid underestimating incidence of HEV infection. Refining our
307 interpretation of hepatitis E serological data, through both improving analytical methods and collecting new
308 longitudinal data across geographies, will be key to expanding the breadth of our understanding of this vaccine-
309 preventable disease.

310 **Acknowledgements**

311 We thank all the participants in the cohort study for taking part in this research. We thank the field and laboratory staff
312 for their dedicated work.

313 Financial Support

314 This work was supported by The Bill and Melinda Gates Foundation (grant number: INV-038404).

315 Conflicts of Interests

316 None.

317 Data and code availability

318 The data and code used in these analyses are available here: <https://github.com/HopkinsIDD/sitakunda-hev>

319 References

- 320 1. **Gerbi GB, et al.** Hepatitis E as a cause of acute jaundice syndrome in Northern Uganda, 2010–2012. *The*
321 *American Journal of Tropical Medicine and Hygiene* 2015; **92**: 411–414.
- 322 2. **Gurley ES, et al.** Estimating the burden of maternal and neonatal deaths associated with jaundice in
323 Bangladesh: possible role of hepatitis E infection. *American Journal of Public Health* 2012; **102**: 2248–2254.
- 324 3. **Paul RC, et al.** Hepatitis E as a cause of adult hospitalization in Bangladesh: Results from an acute jaundice
325 surveillance study in six tertiary hospitals, 2014–2017. *PLOS Neglected Tropical Diseases* 2020; **14**: e0007586.
- 326 4. **Rein DB, et al.** The global burden of hepatitis E virus genotypes 1 and 2 in 2005. *Hepatology* 2012; **55**: 988–
327 997.
- 328 5. **Stanaway JD, et al.** The global burden of viral hepatitis from 1990 to 2013: findings from the Global Burden of
329 Disease Study 2013. *The Lancet* 2016; **388**: 1081–1088.
- 330 6. **Amanya G, et al.** Risk factors, person, place and time characteristics associated with Hepatitis E Virus outbreak
331 in Napak District, Uganda. *BMC Infectious Diseases* 2017; **17**: 451.
- 332 7. **Koyuncu A, et al.** Setting a course for preventing hepatitis E in low and lower-middle-income countries: a
333 systematic review of burden and risk factors. *Open Forum Infectious Diseases* 2021; **8**: ofab178.
- 334 8. **Guthmann J-P, et al.** A large outbreak of hepatitis E among a displaced population in Darfur, Sudan, 2004: the
335 role of water treatment methods. *Clinical Infectious Diseases* 2006; **42**: 1685–1691.
- 336 9. **Zhu F-C, et al.** Efficacy and safety of a recombinant hepatitis E vaccine in healthy adults: a large-scale,
337 randomised, double-blind placebo-controlled, phase 3 trial. *The Lancet* 2010; **376**: 895–902.
- 338 10. **Lynch JA, et al.** Hepatitis E vaccine—illuminating the barriers to use. *PLOS Neglected Tropical Diseases* 2023;
339 **17**: e0010969.
- 340 11. **World Health Organisation** *WHO position paper on hepatitis E vaccines - May 2015*. 2015.
341 (<https://www.who.int/publications-detail-redirect/WER9018-185-200>) Accessed 9 May 2023.

- 342 12. **GBD 2019 Diseases and Injuries Collaborators** Global burden of 369 diseases and injuries in 204 countries
343 and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*
344 2020; **396**: 1204–1222.
- 345 13. **Li P, et al.** The global epidemiology of hepatitis E virus infection: A systematic review and meta-analysis. *Liver*
346 *International: Official Journal of the International Association for the Study of the Liver* 2020; **40**: 1516–1528.
- 347 14. **Muench H.** Derivation of rates from summation data by the catalytic curve. *Journal of the American Statistical*
348 *Association* 1934; **29**: 25–38.
- 349 15. **Kmush BL, et al.** Long-term antibody persistence after Hepatitis E Virus infection and vaccination in Dongtai,
350 China. *Open Forum Infectious Diseases* 2019; **6**: ofz144.
- 351 16. **Kmush BL, et al.** A 10-year immunopersistence study of hepatitis E antibodies in rural Bangladesh. *American*
352 *Journal of Epidemiology* 2018; **187**: 1501–1510.
- 353 17. **Arnold BF, et al.** Enteropathogen antibody dynamics and force of infection among children in low-resource
354 settings: Supplementary Information file 7. *eLife* 2019; **8**: e45594.
- 355 18. **Biswas RSR, et al.** Viral load and genotype of recent hepatitis E virus outbreak in Chittagong, Bangladesh.
356 *Journal of the Scientific Society* 2020; **47**: 144.
- 357 19. **Owada K, et al.** Epidemiological profile of a human Hepatitis E Virus outbreak in 2018, Chattogram,
358 Bangladesh. *Tropical Medicine and Infectious Disease* 2022; **7**: 170.
- 359 20. **Bhuiyan TR, et al.** SARS-CoV-2 seroprevalence before delta variant surge, Chattogram, Bangladesh, March–
360 June 2021. *Emerging Infectious Diseases* 2022; **28**: 429.
- 361 21. **Lumley T.** *survey: analysis of complex survey samples*. 2023.
- 362 22. **WHO/UNICEF.** *Drinking water* | JMP. (<https://washdata.org/monitoring/drinking-water>). Accessed 7 June 2023.
- 363 23. **WHO/UNICEF.** *Sanitation* | JMP. (<https://washdata.org/monitoring/sanitation>). Accessed 7 June 2023.
- 364 24. **Rees EM, et al.** Estimating the duration of seropositivity of human seasonal coronaviruses using seroprevalence
365 studies. *Wellcome Open Research* 2021; **6**: 138.
- 366 25. **Capai L, et al.** Seroprevalence of hepatitis E virus among blood donors on Corsica, France, 2017.
367 *Eurosurveillance* 2020; **25**: 1900336.
- 368 26. **Carpenter B, et al.** Stan: A probabilistic programming language. *Journal of Statistical Software* 2017; **76**: 1–32.
- 369 27. **Stan Development Team.** *RStan: the R interface to Stan*. 2023.
- 370 28. **Goldsmith R, et al.** Enzyme-linked immunosorbent assay for diagnosis of acute sporadic hepatitis E in Egyptian
371 children. *The Lancet* 1992; **339**: 328–331.
- 372 29. **Khuroo MS, et al.** Hepatitis E and long-term antibody status. *The Lancet* 1993; **341**: 1355.
- 373 30. **Vehtari A, et al.** *loo: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models*. 2023.

- 374 31. **Bangladesh Bureau of Statistics**. *Bangladesh Population and Housing Census 2011 - Zila Report -*
375 *Chittagong*. 2015.
- 376 32. **U.S. Census Bureau**. *International Database (IDB)*. (<https://www.census.gov/data-tools/demo/idb>) Accessed 18
377 October 2023.
- 378 33. **Azman AS, et al**. Hepatitis E in Bangladesh: Insights From a National Serosurvey. *The Journal of Infectious*
379 *Diseases* 2021; **224**: S805–S812.
- 380 34. **Labrique AB, et al**. Epidemiology and risk factors of incident Hepatitis E Virus infections in rural Bangladesh.
381 *American Journal of Epidemiology* 2010; **172**: 952–961.
- 382 35. **Labrique AB, et al**. Population seroprevalence of hepatitis E virus antibodies in rural Bangladesh. *The*
383 *American Journal of Tropical Medicine and Hygiene* 2009; **81**: 875–881.
- 384 36. **Gurley ES, et al**. Outbreak of Hepatitis E in Urban Bangladesh Resulting in Maternal and Perinatal Mortality.
385 *Clinical Infectious Diseases* 2014; **59**: 658–665.
- 386 37. **Zhang J, et al**. Long-term efficacy of a hepatitis E vaccine. *New England Journal of Medicine* 2015; **372**: 914–
387 922.
- 388 38. **Bendall R, et al**. A comparison of two commercially available anti-HEV IgG kits and a re-evaluation of anti-HEV
389 IgG seroprevalence data in developed countries. *Journal of Medical Virology* 2010; **82**: 799–805.
- 390 39. **Yan Q, et al**. Comparison of two diagnostic reagents to detect anti-hepatitis E virus IgG antibodies. *Chin J*
391 *Zoonoses* 2008; **24**: 1087–1089.
- 392 40. **Hoang TN, et al**. A tightly clustered hepatitis E virus genotype 1a is associated with endemic and outbreak
393 infections in Bangladesh. *PLOS ONE* 2021; **16**: e0255054.
- 394 41. **Haider N, et al**. Serological evidence of hepatitis E virus infection in pigs and jaundice among pig handlers in
395 Bangladesh. *Zoonoses and Public Health* 2017; **64**: 572–577.