Hepatitis E Virus Seroprevalence among Adults, Germany

Mirko S. Faber,1 Jürgen J. Wenzel,1 Wolfgang Jilg, Michael Thamm, Michael Höhle, and Klaus Stark

We assessed hepatitis E virus (HEV) antibody seroprevalence in a sample of the adult population in Germany. Overall HEV IgG prevalence was 16.8% (95% CI 15.6%–17.9%) and increased with age, leveling off at >60 years of age. HEV is endemic in Germany, and the lifetime risk for exposure is high.

In industrialized countries, hepatitis E virus (HEV) has long been regarded as a rare imported infection. However, sporadic cases without travel to disease-endemic areas and causing genotype 3 are being increasingly reported (1,2). Epidemiologic and molecular studies have implicated undercooked pork and wild boar products as a source of HEV infection (3–5). An unexpectedly high prevalence of HEV-specific antibodies, e.g., among blood donors, has been shown by several studies in Europe and the United States (6–11).

In Germany, the number of notified hepatitis E cases rose from <50 annually in 2001–2003 to 238 in 2011 (incidence 0.3/100,000 population); the proportion of autochthonous cases increased from 30%–40% to 78%. We conducted a study to determine HEV seroprevalence in Germany’s adult population and associations with sociodemographic characteristics by using an assay highly sensitive for HEV genotype 3.

The Study

We assessed HEV seroprevalence in a large subsample (n = 4,422) of the 2008–2011 German Health Examination Survey for Adults (Deutscher Erwachsenen Gesundheitssurvey; www.degs-studie.de), a 2-stage national probability sample that assessed the health status and underlying assumptions are described in the online Technical Appendix (wwwnc.cdc.gov/EID/pdfs/11-1756-Techapp.pdf).

The 4,352 persons who were included in the analysis were from 108 communities of all federal states in Germany (Table 1). Weighted prevalence of HEV IgG was 16.8% (95% CI 15.6%–17.9%); prevalence ranged from 6.1% (95% CI 4.5%–7.8%) in the 18–34-year age group to >20% in the >50-year groups, with a maximum of 46.6% in the sample vs. 8.7% in the total adult population.

Serum samples were screened for HEV IgG by using the recomLine HEV-IgG/IgM immunoassay (Mikrogen, Neuried, Germany). The assay is based on recombinantly expressed antigens of genotypes 1 and 3 of open reading frames 2 and 3. According to the manufacturer’s and our data (J.J. Wenzel et al., unpub. data), the test is 97%–100% sensitive for detecting acute or previous HEV infections. Test strips were scanned with the semiautomatic recomScan software (Mikrogen). The intensity of 3 quality assurance and other bands was determined by densitometrical detection algorithms. Each antigen band with an intensity greater or equal to the cutoff was assigned a point value. The final results were classified into 3 categories: no antibodies detectable (negative), test inconclusive (borderline), and antibodies detectable (positive). Persons whose test results were borderline (n = 70) were excluded from further analysis.

We poststratified the remaining survey population (n = 4,352) by age group and location of residence (16 states) to account for per protocol oversampling in eastern Germany and to restore the distribution of age groups to match the distribution in the total population. Weighted seroprevalence estimates were calculated by using survey-weighted logistic regression. Associations between demographic characteristics and seropositivity were analyzed by using adjusted Wald test p values. We also estimated mean annual incidence of HEV seroconversion from the seroprevalence data by using a catalytic model with age-constant force of infection, similar to that of Faramawi et al. (12). Detailed methods and underlying assumptions are described in the online Technical Appendix (wwwnc.cdc.gov/EID/pdfs/11-1756-Techapp.pdf).

The 4,352 persons who were included in the analysis reflects the total adult population with respect to age, sex, and geographic region, but persons with migration background are underrepresented (non-German citizenship 4.6% in the sample vs. 8.7% in the total adult population).

Author affiliations: Robert Koch Institute, Berlin, Germany (M.S. Faber, M. Thamm, M. Höhle, K. Stark); and University of Regensburg, Regensburg, Germany (J.J. Wenzel, W. Jilg)

DOI: http://dx.doi.org/10.3201/eid1810.111756

1These authors contributed equally to this article.
Hepatitis E Virus Seroprevalence, Germany

Mean annual incidence of HEV seroconversion estimated from the catalytic model was 3.9 (95% CI 3.6%–4.2%) per 1,000 population.

Conclusions

We found an overall HEV seroprevalence of 16.8% among adults in Germany; seroprevalence increased with age but was not dependent on sex or location of residence. Similarly high seroprevalence was found in blood donors in Denmark (20.6%), southwestern England (16%), and the United States (18%) (6,7,9). Estimates from Switzerland and the Netherlands, on the other hand, are considerably lower (10,11). Reasons for these differences could be effects of sample selection, different lifetime exposures (e.g., to foods that may serve as transmission vehicles), or use of different test systems with varying sensitivity. Mansuy et al. recently reported 53% prevalence of HEV antibodies in blood donors in southwestern France (13), a figure considerably higher than the 17% prevalence reported earlier for the same geographic region, when a different test system was used (8). The assay applied in our study was designed to also detect previous infections with HEV genotype 3; therefore, it is likely to be more sensitive than assays used in previous studies.

Our data show that HEV exposure is common among the general adult population in Germany, which is consistent with increasing evidence for pigs as a reservoir for foodborne transmission of HEV in industrialized countries. HEV seroprevalence is high in domestic pig herds in Germany and other countries; closely related HEV strains were found in pig livers on retail sale and in autochthonous cases of hepatitis E, and HEV seroprevalence was higher in persons with occupational exposure to pigs than in control groups (4,10,14,15).

Our data and other studies (11–14) have shown no significant difference in seroprevalence between sexes, despite an assumed higher frequency of alimentary or occupational exposure and the higher incidence of clinical cases among men. This finding may indicate sex-specific differences in disease development or application of laboratory testing or that foods frequently consumed by both sexes play a substantial role as vehicles for transmission. These results also may highlight the lack of evidence for 1 main risk factor or food vehicle (14).

The strong association between age and HEV seroprevalence in our study most likely reflects cumulative lifetime exposure to the virus. However, HEV seroprevalence remains relatively stable from 18 to 34 years of age, which could indicate a birth cohort effect resulting from a decrease in the overall risk over the past few decades, as was reported for Denmark (6). The leveling
off of the seroprevalence above age 60 years could be caused by a loss of antibodies in the elderly.

We found a striking difference between the estimated annual incidence of seroconversion and the relatively low incidence of notified disease in Germany. Besides underdiagnosis, a possible explanation is a high proportion of asymptomatic infections.

The survey response rate is similar to those achieved in other comprehensive national health examination surveys in Europe. In the statistical analysis, we corrected for differences in the distribution of age and place of residence between the sample and the general population. However, foreign-born persons are underrepresented in the survey and location of residence, Germany, 2008–2011.*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Seroprevalence, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>16.8 (15.6–17.9)</td>
</tr>
<tr>
<td>Age group, y</td>
<td></td>
</tr>
<tr>
<td>18–19</td>
<td>5.3 (2.5–11.1)</td>
</tr>
<tr>
<td>20–29</td>
<td>6.6 (4.8–9.2)</td>
</tr>
<tr>
<td>30–39</td>
<td>8.0 (5.8–10.8)</td>
</tr>
<tr>
<td>40–49</td>
<td>16.9 (14.3–19.9)</td>
</tr>
<tr>
<td>50–59</td>
<td>22.6 (19.7–25.7)</td>
</tr>
<tr>
<td>60–69</td>
<td>26.1 (23.0–29.4)</td>
</tr>
<tr>
<td>70–79</td>
<td>23.8 (20.4–27.5)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>16.8 (15.2–18.5)</td>
</tr>
<tr>
<td>F</td>
<td>16.7 (15.2–18.5)</td>
</tr>
<tr>
<td>Place of residence (north–south)</td>
<td></td>
</tr>
<tr>
<td>Northern states</td>
<td>15.5 (13.5–17.6)</td>
</tr>
<tr>
<td>Middle states</td>
<td>16.6 (14.8–18.6)</td>
</tr>
<tr>
<td>Southern states</td>
<td>17.9 (15.8–20.1)</td>
</tr>
<tr>
<td>Place of residence (east-west)</td>
<td></td>
</tr>
<tr>
<td>Western states</td>
<td>16.6 (15.2–18.0)</td>
</tr>
<tr>
<td>Eastern states</td>
<td>17.6 (15.6–19.7)</td>
</tr>
<tr>
<td>Population of municipality</td>
<td></td>
</tr>
<tr>
<td>&lt;5,000</td>
<td>17.9 (15.3–20.8)</td>
</tr>
<tr>
<td>5,000–&lt;20,000</td>
<td>17.5 (15.2–20.1)</td>
</tr>
<tr>
<td>20,000–&lt;100,000</td>
<td>14.4 (12.4–16.6)</td>
</tr>
<tr>
<td>&gt;100,000</td>
<td>17.7 (15.7–20.0)</td>
</tr>
</tbody>
</table>

*n = 4,362.

Acknowledgments

We thank Monika Erl and Bianka Ehrlich for expert technical assistance and Heribert Stolzenberg and Panagiotis Kamtsiuris for critical review of the manuscript.

No external funding was received for this study. The study was approved by the ethical review board of the Charité University Hospital, Berlin.

Dr Faber is a medical epidemiologist at the Robert Koch Institute, Department for Infectious Disease Epidemiology (Gastrointestinal Infections, Zoonoses and Tropical Infections Unit).

References


Address for correspondence: Mirko S. Faber, Department for Infectious Disease Epidemiology, Robert Koch Institute, DGZ-Ring 1, 13086 Berlin, Germany; email: faberm@rki.de
Hepatitis E Virus Seroprevalence among Adults, Germany

Technical Appendix

Detailed Methods

Description of the catalytic model

The incidence in the manuscript was computed using a so called simple catalytic epidemic model (Griffiths 1974; Farrington 2005), where the force of infection (FOI) is assumed to be time constant. The dependent variable in the fitted models was a binary variable indicating whether a person in the study had seroconverted or not. Catalytic models assume that infection induces life-long immunity and does not affect the mortality rate of infected individuals. A consequence of the constant FOI is that the population is assumed to be homogeneous with respect to both susceptibility and exposure to infection. Furthermore, infection is assumed to be in equilibrium state, i.e. the level of incidence is assumed to remain constant in time.

Notation

Denote the available data \( \{(y_i,a_i), i=1,...,n\} \) where \( y_i \) is a binary variable indicating if the i’th individual has seroconverted (0 = no, 1 = yes) and \( a_i \) is the age of the individual in years (i.e. taken as a continuous variable). Let \( p(a) \) be the probability that an individual of age \( a \) has sero-converted. Inference about the parameters in a parametric model for \( p(a) \) can now be performed using the binomial likelihood:

\[
L = \prod_{i=1}^{n} p(a_i)^{y_i} (1 - p(a_i))^{1-y_i}.
\]

In the constant FOI model, i.e. \( \lambda(a) = \alpha \) for \( a \geq 0 \), the probability to have seroconverted at age \( a \) is given by \( p(a) = 1 - \exp(-\alpha a) \). One can show (see, e.g., Becker 1989 or Farrington 2005) that the desired estimation problem for \( \alpha \) can be reduced to the fitting of a generalized linear model with complementary log-log (cloglog) link function having the following linear predictor:

\[
\eta(a) = \log(-\log(1 - p(a))) = \log(\alpha) + \log(a).
\]
This model can be fitted with any GLM software (e.g., function glm in Stata or R) by specifying a binomial model with cloglog link function and using log(a) as offset in the linear predictor. The natural exponent of the intercept estimate in such a model is the desired estimate \( \alpha \). The annual incidence can now be computed as

\[
I = \frac{p(a + 1) - p(a)}{p(a)} = 1 - \exp(-\alpha).
\]

In its basic form, this is the model used by Faramawi et al. (2011) to compute the annual incidence. Confidence intervals for \( I \) are easily obtained by transforming the confidence interval of the intercept in the GLM by the above equation for \( I \). We chose this simple constant FOI in order to obtain a nation-wide estimate of the annual incidence which allows for comparison with the Faramawi et al. estimates for the US.

**Results**

Performing the above calculations for our \( n = 4,352 \) individuals we obtain \( \hat{I} = 0.00398 \), i.e., the annual incidence is 398 per 100,000 population with a 95% CI of 372–428. To address post-stratification in the analysis, we additional fitted the above GLM by weighting each observation according to its specific post-stratification sampling weight (e.g. function svyglm in Stata or R). The (survey-weighted) estimated annual incidence is now 392 per 100,000 population (95% CI 364–423). We report this weighted estimate of the annual incidence in the manuscript together with the associated confidence intervals.

**Model Checking**

Graphical analysis of the residuals from a model with binary response is difficult due to the extreme discrete nature of the problem. Furthermore, such model checking is further complicated by the complex survey setup of our sample. Instead, we perform an alternative examination of the model fit. As a first qualitative assessment, we decided to investigate the models point-wise predictive performance. Based on the asymptotic normality of the GLM estimate we sampled 999 FOI estimates from the normal distribution with mean \( \hat{\alpha} \) and variance equal to the estimated variance of \( \hat{\alpha} \). For each FOI obtained from this sampling we then
• Calculated the model predicted probability \( \hat{p}(a_i) \) for each individual \( i = 1,\ldots,4352 \) in the data and sampled a Bernoulli variable using this probability for each individual.

• Given the above Bernoulli realizations we could then for each year of age obtain a new raw seroprevalence estimate respecting the sampling weights (i.e., by using the function svymean).

The Technical Appendix Figure shows point-wise 95% prediction intervals for the seroprevalence of age (in years) based on these 999 parametric bootstrap samples. For comparison, the original survey weighted estimates are indicated as black dots with size proportional to the number of observations available at that age. We observe that the simple catalytic model obtains a good fit with only few points outside the prediction bands. Also, the Pearson goodness of fit test for the simple catalytic model does not reject the null hypothesis (\( p = 0.88 \)).

![Technical Appendix Figure](image_url)

Technical Appendix Figure. Survey-weighted seroprevalence for each year illustrated as black dots. Each dot is shown proportional in size to the actual sample size (n) at that age. 95% point-wise prediction intervals for the simple catalytic are shown in grey. Also shown are the model fitted proportions for the simple catalytic model and two additional models with extra flexibility for age.

In the figure, the black dotted line shows the estimated \( p(a) \) of the simple catalytic model. As a sensitivity analysis the red line similarly shows \( p(a) \) in Weibull model for the FOI. This model obtains a better fit with \( \gamma \) being significant indicating that the FOI, and hence the annual
incidence, is age specific. However, age specific incidence rates would be harder to interpret and report, especially because our aim was to calculate an overall incidence rate. Another reason is that even this model is not fully sufficient to address all aspects of the data: the green line shows the fit of a fully flexible survey-weighted kernel smoother for \( p(a) \). In concordance with Figure 1 of the manuscript we here observe a slight decrease in the seroprevalence at the high ages, which is mentioned in the manuscript discussion. Reporting age-specific annual incidence rate based on such a flexible model would require a table containing a number for each year of age, which is not really useful for our purpose.

Within the cloglog GLM framework it is possible to allow for heterogeneity in the FOI by adjusting for additional variables than age. We investigated additional dependence of sex and residence in the linear predictor, but none of these variables turned out to be significant at the 5% significance level.

Thus, our reported annual incidence estimate remains a nicely interpretable and communicable result which allows for comparison with Faramawi et al., while the above model checking indicates that the assumptions of the constant rate model are reasonable.

**References**