**Bimodal distribution and set point HBV DNA**

**viral loads in chronic infection: retrospective**

**analysis of cohorts from the UK and South Africa**

**Supplementary Methods**

***Clinical datasets***

**(i) UK HBV dataset (n=351 patients, 1483 timepoints)**

UK CHB data were collected from electronic records at Oxford University Hospital NHS Foundation Trust, as part of the National Institute of Health Research Health Informatics Collaborative (NIHR-HIC, [www.hic.nihr.ac.uk](http://www.hic.nihr.ac.uk)) as previously described (1), collecting data for 371 individuals. Chronic HBV infection was defined as positive HBsAg on ≥2 occasions ≥6 months apart. We also collected age, sex, self-reported ethnicity, and HBeAg status, generating a complete data set for 351 individuals. For analysis of longitudinal VL data, we used only VL measurements prior to commencing antiviral treatment, and included patients with ≥2 VL measurements (n= 299 patients).

HBV and HCV VL are conventionally reported in IU/ml, based on international standards set by the WHO, but to make direct comparisons between VL in different infections, we also converted data into copies/ml (1 IU = 5.4 copies/ml for HBV (2) and 2.7 copies/ml for HCV (3)).

**(ii) South Africa dataset (n=6506 patients; 9578 timepoints)**

We collected all HBV VL data from the South African National Health Laboratory Service (NHLS) recorded between 1st January 2015 and 31st December 2018. For the purposes of analysis and to reduce possible errors in raw data, all VL measurements below the limit of detection were excluded from analysis with the assumption they were on anti-viral treatment, either for HBV or for co-infection with HIV, and all those above the standard laboratory limit of quantification were designated 1.7x108 IU/ml. For analysis of longitudinal data, only those patients with ≥2 detectable VL measurements were included (n=874). HBV VL samples used in median calculation and frequency distribution for South African data included multiple longitudinal samples for single individuals. Other data (HBeAg status, HIV status, treatment data) were not available for this cohort.

**(iii) HCV and HIV cohorts**

UK HCV data were collected from the same source as the UK HBV data. Baseline HCV viral loads were collected for all patients within the Oxford NIHR-HIC database prior to commencing antiviral treatment between 2006-2018.

HIV data were obtained from an established UK database of HIV seroconverters between 1985-2014 developed as part of the BEEHIVE collaboration (n=1581). HIV viral loads were measured using COBAS AmpliPrep/COBAS TaqMan HIV-1 Test, v2.0 on samples collected between 6-24 months after infection. The average VL for each patient over this time period was used as their set point viral load (SPVL) (4).

**Statistical analysis**

We used Graphpad Prism v.8.2.1. for analysis of viral load distributions, calculation of skewness, and univariate analysis of patient parameters associated with HBV VL (Mann Whitney U test and Kruskall Wallis test). Bin sizes of 0.5 log10 IU/ml were used for the histograms. Longitudinal analysis for the UK cohort included all patients with ≥2 VL measurements recorded off treatment. We used R package to assess within and between patient variability. We defined total variation, within-individual and between-individual variation according to analysis of variance (ANOVA). Specifically, the calculations are as follows:

Where denotes the number of patients, represents the number of data points for patient , denotes the viral load of patient at time point , is the mean of viral loads of all data points of a cohort, is the mean of viral loads of patient .

**Ethics Approval**

Data collection for the UK cohort was approved by the NRES Committee South Central-Oxford C (ref: 15/SC/0523). South African data collection was approved by the Health Sciences Research Ethics Committee at the University of the Free State; reference UFS-HSD2019/0044/2603.

**Data Availability**

Full metadata for this study are available on-line on Figshare.

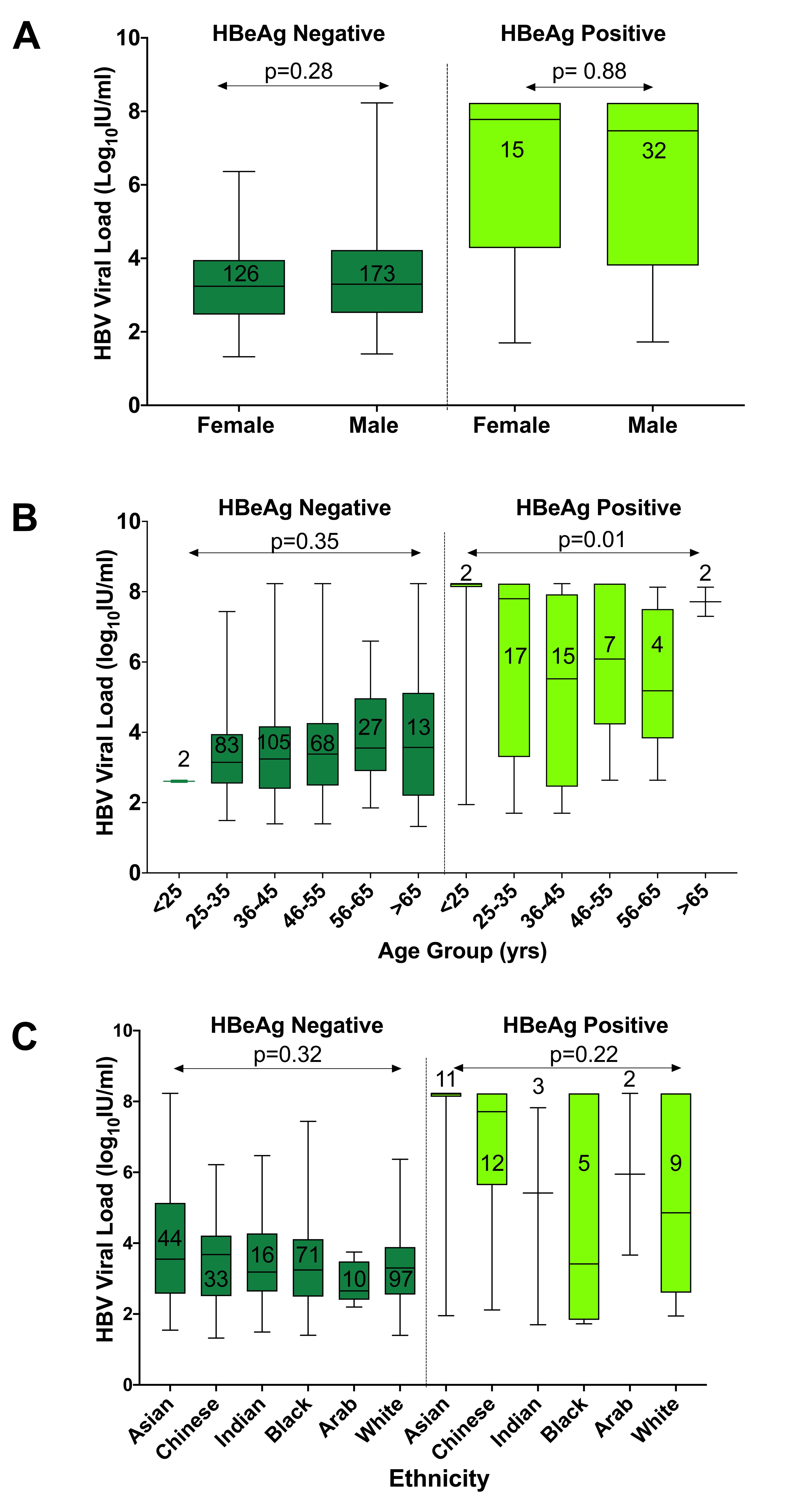
Temporary URL prior to publication: [https://figshare.com/s/cd8b046346250dd349b0](https://figshare.com/s/cd8b046346250dd349b0" \t "_blank)

This will be converted to a permanent DOI on acceptance of the data for publication, 10.6084/m9.figshare.11365082.

**Supplementary Results**

The average age of our UK HBV cohort was 42 years, with the dominant ethnic groups being White, Black and Asian (34%, 24% and 17% respectively). The majority of our cohort was HBeAg negative (81%) and male (56%).

**Supplementary Figure 1: Relationship between HBV viral load and other patient characteristics, grouped according to HBeAg positive or negative status in UK cohort.** (A) Sex; (B) Age; (C) Ethnicity. Ethnicity was self-reported using standardised codes. Ethnic groups containing <10 individuals were excluded. The Asian group includes patients from Asia other than the Indian subcontinent and China/Hong Kong. In each panel, box and whiskers show median, IQR and range for baseline HBV VL (IU/ml). Numbers in boxes correspond to patient numbers. P-values by Mann-Whitney test (panel A) and Kruskall-Wallis (panels B and C).

****

**References**

1. Downs LO, Smith DA, Lumley SF, Patel M, McNaughton AL, Mokaya J, et al. Electronic Health Informatics Data To Describe Clearance Dynamics of Hepatitis B Surface Antigen (HBsAg) and e Antigen (HBeAg) in Chronic Hepatitis B Virus Infection. MBio. 2019;10(3).

2. Valsamakis A. Molecular testing in the diagnosis and management of chronic hepatitis B. Clin Microbiol Rev. 2007; 20(3):426–39, table of contents.

3. Pawlotsky JM. Use and interpretation of virological tests for hepatitis C. In: Hepatology. 2002;36(5 Suppl 1):S65-73

4. Blanquart F, Wymant C, Cornelissen M, Gall A, Bakker M, Bezemer D, et al. Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. PLoS Biol. 2017;15(6):e2001855.