Data from an electronic health informatics pipeline to describe clearance dynamics of Hepatitis B surface antigen (HBsAg) and e-Antigen (HBeAg) in chronic HBV infection

Louise O Downs[1,2*], David Smith[2,3*], Sheila F Lumley[1,2], Meha Patel[1], Anna L McNaughton[2], Jolynne Mokaya[2], M Azim Ansari[2], Hizni Salih[4], Kinga A Várnai[3], Oliver Freeman[4], Sarah Cripps[5], Jane Phillips[6], Jane Collier[6], Kerrie Woods[3], Keith Channon[3], Jim Davies[4], Eleanor Barnes[2,3,6], Katie Jeffery[1], Philippa C Matthews[1,2,3#]

[*] These two authors contributed equally

[1] Department of Infectious Diseases and Microbiology, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK

[2] Nuffield Department of Medicine, University of Oxford, Medawar Building for Pathogen Research, South Parks Road, Oxford OX1 3SY, UK

[3] National Institutes of Health Research Health Informatics Collaborative, NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford OX3 9DU

[4] Oxford NIHR Biomedical Research Centre Clinical Informatics Group, Big Data Institute, University of Oxford, Li Ka Shing Centre for Health Information and Discovery, Old Road Campus, Oxford OX3 7LF, UK

[5] Pharmacy Department, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK

[6] Department of Hepatology, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK

[#] **CORRESPONDING AUTHOR**

Email: philippa.matthews@ndm.ox.ac.uk

Address: Medawar Building for Pathogen Research, South Parks Road, Oxford OX1 3SY, UK

Telephone: 0044 1865 271973

KEY WORDS: HBV, HBsAg, HBeAg, surface antigen, clearance, cure, dynamics, treatment, prognosis, monitoring, biomarkers

CONFLICT OF INTEREST: Nil

FINANCIAL SUPPORT:

LD is funded by the National Institute for Health Research (NIHR). PCM is funded by the Wellcome Trust, grant number 110110. EB is funded by the Medical Research Council UK, the Oxford NIHR Biomedical Research Centre and is an NIHR Senior Investigator. The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

AUTHORS' CONTRIBUTIONS:

- Conceived the study: KJ, PCM
- Assimilated/analysed clinical records: LD, SL, JM, MP, SC, KJ
- Established the BRC informatics working group: KC
- Developed the health informatics (HIC) pipeline: DS, HS, JD, OF, KV, KW
- Lead for viral hepatitis health informatics: EB
- Analysed the data: DS, ALM, MAA, SL, JM, PCM
- Literature review: LD, JM
- Wrote the manuscript: LD, DS, PCM
- Edited and approved the final manuscript: all authors

ACKNOWLEDGEMENTS

This work has been facilitated by infrastructure development at Oxford Biomedical Research Centre (BRC) funded by the NIHR Health Informatics Collaborative (HIC).

An earlier version of this dataset was presented as a poster at the EASL International Liver Congress meeting, Paris 2018 (https://f1000research.com/posters/7-917).

.

ABBREVIATIONS

CHB	Chronic Hepatitis B virus infection
EPR	Electronic patient record
HBeAg	Hepatitis B ‘e’ antigen
HBsAg	Hepatitis B surface antigen
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HIC	Health Informatics Collaborative
HIV	Human immunodeficiency virus
LFT	Liver function tests
LIMS	Laboratory information management system
NA	Nucleos(t)ide analogues
NIHR	National Institute for Health Research
PEG-IFNα	Pegylated interferon alpha 2
TDF	Tenofovir Disoproxil Fumarate
RBV	Ribavirin

ABSTRACT

HBsAg and HBeAg have gained traction as biomarkers of control and clearance during chronic hepatitis B virus infection (CHB). Improved understanding of clearance correlates of these proteins could help inform improvements in patient-stratified care and advance insights into underlying mechanisms of disease control, thus underpinning new cure strategies. We collected electronic clinical data via an electronic pipeline supported by the National Institute for Health Research Health Informatics Collaborative (NIHR-HIC), adopting an unbiased approach to generating a robust longitudinal dataset for adults testing HBsAg-positive from a large UK teaching hospital over a six year period (2011-2016 inclusive). From 553 individuals with CHB, longitudinal data were available for 319, representing >107,000 weeks of clinical follow-up. Among these 319 individuals, 13 (4%) cleared HBsAg completely. Among these 13, HBsAg clearance rate was similar in individuals on nucleos(t)ide analogue (NA) therapy (n=4 (31%)), median clearance time 150 weeks) vs those not on NA therapy (n=9 (69%), median clearance time 157 weeks). Those who cleared HBsAg were significantly older, and less likely to be on NA therapy compared to non-clearers (p=0.003 and p=0.001, respectively). Chinese ethnicity was associated with HBeAg positivity (p=0.025). HBeAg clearance occurred both on NA therapy (n=24, median time 49 weeks) and off NA therapy (n=19, median time 52 weeks). Improved insights into the dynamics of these biomarkers can underpin better prognostication and patient-stratified care. Our systematised approach to data collection paves the way for scaling up efforts to harness clinical data to address research questions and underpin improvements in clinical care.

IMPORTANCE

Advances in the diagnosis, monitoring and treatment of hepatitis B virus (HBV) infection are urgently required if we are to meet international targets for elimination by the year 2030. Here we demonstrate how routine clinical data can be harnessed through an unbiased electronic pipeline, showcasing the significant potential for amassing large clinical datasets that can help to inform advances in patient care, and provide insights that may help to inform new cure strategies. Our cohort from a large UK hospital includes adults from diverse ethnic groups that have previously been under-represented in the literature. Tracking two protein biomarkers that are used to monitor

chronic HBV infection, we provide new insights into the timelines of HBV clearance, both on and off treatment. These results contribute to improvements in individualised clinical care and may provide important clues into the immune events that underpin disease control.

INTRODUCTION

Chronic HBV (CHB) infection is defined as detectable HBsAg (>20 IU/ml) at ≥2 timepoints ≥6 months apart. Disease activity and treatment response in individuals with CHB infection are most commonly monitored by quantification of HBV DNA viral load (1). However, viral load measurement is expensive, and not universally available, viral DNA levels can fluctuate over time, and quantification can be inaccurate at low levels. Reproducible, automated quantification of other biomarkers such as hepatitis B surface antigen (HBsAg) and/or e-antigen (HBeAg) are therefore attractive biomarkers for use instead of, or alongside, HBV DNA monitoring.

In the context of CHB infection, HBV cccDNA persists as an intranuclear 'mini-chromosome' within infected hepatocytes (Fig 1A). HBsAg is produced in excess, from translation of both from the cccDNA reservoir, and from pre-genomic mRNA (2). In a small proportion of cases, HBsAg becomes undetectable over time, suggesting that the cccDNA reservoir is down-regulated, diminished or lost entirely. In the HBV cure field, specific terminology has been adopted to reflect the difference between complete loss of all cccDNA ('sterilising cure') vs suppression or dilution of cccDNA to the extent that HBsAg can no longer be detected ('functional cure') (3, 4). In practical terms, there is no current way to differentiate between these two outcomes. However, the theoretical distinction is an important one, as sterilising cure reflects complete and permanent loss of HBV from the host, while in the setting of functional cure, there is long-term potential for relapse to occur (best recognised in the setting of immunosuppression (5, 6).

Nucleos(t)ide analogues (NA's) inhibit HBV reverse transcriptase, leading to loss of HBV DNA from the serum, but have no direct effect on cccDNA (3). Thus, HBsAg production can continue unchecked from the cccDNA reservoir, and viral replication frequently returns on cessation of treatment (fig 1B) (3). For this reason, most guidelines currently recommend long term NA treatment (1). Immune responses (either arising naturally or driven by immunotherapy such as interferon) can lead to downregulation or loss of cccDNA to the extent that neither HBsAg nor HBV DNA can be detected in the serum (fig 1C) (4). The long term goal of new immunotherapeutic

approaches will be to elicit sterilising cure such that cccDNA is removed with no long-term risk of relapse (7).

HBsAg levels are typically highest in the earlier phases of infection and in HBeAg-positive individuals, frequently correlate with HBV DNA levels in CHB infection, and are associated with risk of subsequent reactivation (8). HBsAg may be a quantifiable risk factor for development of hepatocellular carcinoma (HCC) and chronic liver disease (9), although the relationship is not well defined: in some studies, higher HBsAg levels are associated with lower levels of fibrosis (10–12), while in others, lower baseline HBsAg levels are associated with reduced risk of both cirrhosis and HCC (13). HBsAg levels have also been used to classify individuals into those with inactive carriage (HBV DNA <2000 IU/ml and normal ALT (14, 15)) versus active CHB (with higher viral loads and elevated risks of inflammatory liver disease, fibrosis and cirrhosis (16–19)). HBsAg elimination is widely regarded as a marker of immunological clearance (which may be regarded as 'functional cure').

HBeAg-positivity is associated with high viral loads and is therefore a marker of infectivity. Loss of HBeAg is usually associated with production of anti-HBe antibody (a marker of immune-mediated control), and typically associated with lower viral loads. However, although these broad patterns have been described, further efforts are required to elucidate and interpret the dynamics of HBsAg and HBeAg, with the potential to develop insights into the timing and patterns of immunological clearance, and to improve patient-stratified clinical management.

A recent systematic review and meta-analysis has collated literature on HBsAg clearance, with a primary focus on untreated populations (20). This identified 34 studies, but only 14 of these reported ≥2 HBsAg measurements over time, and all but two were in Asia. To ensure we had adequately reviewed the relevant existing evidence on this topic, we also undertook an independent literature review (summarised in Suppl Table 1). We initially identified 43 studies reporting dynamics of HBsAg loss in CHB infection. We excluded studies prior to 2008, those reporting only one HBsAg measurement, and those without an annual or cumulative HBsAg

clearance rate, leaving nine relevant studies. As for the meta-analysis, the majority (8/9) were in Asian populations (21–28), with the remaining one based in New Zealand (29). The reported clearance rate of HBsAg ranged from 0.15% per year (27) to 2.7% per year (24) with a maximum cumulative clearance of 3.5% (21). Older age was associated with clearance in two cohorts (23, 29). The role of treatment in clearance is inconsistent, with NA treatment associated with clearance in some cohorts (21, 25) but not in others (26).

HBsAg levels can be used to determine treatment response, although this has been more reliably reported for PEG-IFN2α treatment than for NAs (30, 31), as it implies reduction or removal of the cccDNA reservoir (Fig 1C). Current UK guidelines recommend quantitative HBsAg and HBeAg measurement before starting treatment and at weeks 12, 24 and 48 during treatment, followed by 6 monthly measurement during long term therapy (32). European Association for the Study of the Liver (EASL) guidelines recommend quantitative HBsAg measurement annually in treated patients if HBV DNA is undetectable, as well as using HBsAg levels to inform the decision to stop treatment (1). EASL guidelines also recommend HBeAg measurement as part of the initial clinical assessment, and list HBeAg loss as one of the serological responses to treatment, but do not specify a frequency for follow-up testing (1).

International targets arising from the United Nations ‘sustainable development goals’ have set a challenge for elimination of CHB infection as a public health threat by the year 2030 (33). Recognising the multi-lateral approaches that will be required to reach this ambitious goal, we here focus on two inter-related aims:

i. We set out to showcase how longitudinal data for individuals with CHB can be collected through an unbiased electronic pipeline that collates, cleans and anonymises routinely-collected electronic clinical data, in this case driven by infrastructure supported by the UK National Institute for Health Research (NIHR) Health Informatics Collaborative (HIC); (www.hic.nihr.ac.uk). The aim is to harness clinical data to drive research and quality improvements in diagnostics, monitoring and therapy of viral hepatitis, and to underpin new questions for basic science. Through developing and testing this system, we have devised

an approach that can be rolled-out to incorporate other centres, with substantial gains predicted through the power of large datasets.

ii. We analysed data for HBV sourced from a tertiary referral UK teaching hospital, in order to develop better insights into patterns of HBsAg and HBeAg clearance. Through the application of an unbiased approach (agnostic to treatment, clinical stage of disease, other biomarkers, or genotype of infection), we aim to develop a clear picture of the dynamics of clearance. Identifying demographic or clinical characteristics that predict specific disease outcomes, provides opportunity for the investigation of immunological correlates of control and clearance.

Collectively, this enterprise provides proof-of-principle for the systematic use of electronic clinical data in informing studies of viral hepatitis, as well as shedding new light on the dynamics of clearance of HBsAg and HBeAg.

RESULTS

Description of a clinical cohort of chronic HBV infection

We identified 553 individuals who tested HBsAg-positive during the six-year period 2011-2016, inclusive. Of these, 319 met inclusion criteria for further analysis (as shown in Table 1; Fig 2). Characteristics of the cohort are summarised in Table 2 and the complete metadata for these 319 CHB patients is available as a supporting data file (Suppl Table 2). We collected longitudinal data for a total of 107,702 person-weeks (range 61-702 weeks, mean 338 weeks (6.5 years) of follow-up per individual, IQR 174-487). The median age at first HBsAg test was 34 years (IQR 29 - 43, range 10 – 71), and males accounted for 191/319 (60%) of cases. HIV co-infection was documented in 9 individuals (2.8%), although we cannot exclude the possibility that the true prevalence of HIV-coinfection was higher due to a proportion of individuals who did not have a recent HIV test result.

Frequency of HBsAg clearance

Exemplar patterns of HBsAg clearance are illustrated in Fig 3 (as per definitions in Table 1). Using the most stringent definition of HBsAg clearance, we documented complete clearance in 13/319

(4.1%) individuals (for full details see Suppl Table 2 and clearance trajectories shown in Suppl Fig 1). The HBsAg clearance rate for this cohort was 0.6% per year. In only 2/13 cases could we estimate the likely duration of infection prior to clearance, one individual who had been vertically infected (HBS-145) and one with iatrogenic infection related to a blood transfusion in childhood (HBS-113). These individuals were both infected for approximately 25 years before clearing HBsAg.

We classified an additional 27/319 (8.5%) individuals as 'potential clearers' on the grounds of HBsAg trends consistently declining towards clearance (criteria in Table 1; clearance curves shown in Suppl Fig 2). These represent a more heterogenous group, but the clearance trajectory in all cases suggests that they would meet the more stringent clearance criteria if prospective surveillance were to be continued. In contrast, HBsAg curves for non-clearers are shown in Suppl Fig 3.

Characteristics of individuals with HBsAg clearance or potential clearance

Adults classified as completely or potentially clearing HBsAg were significantly older than non-clearers (median age 40 vs 34 years; p=0.003; Table 2; Suppl Fig 4). There was no difference in sex or ethnic origin between individuals in different HBsAg clearance categories (Table 2). The majority of those who completely cleared HBsAg were HBeAg-negative throughout the period of observation (10/13, 77%). Among the remaining three with detectable HBeAg, two of these lost HBeAg prior to clearing HBsAg (HBS-197 and HBS-223), while one (HBS-195) cleared HBsAg and HBeAg together (Suppl Fig 1). In three cases (HBS-113, HBS-145 and HBS-195), HBV DNA was cleared at the same time as HBsAg; however, in the other ten individuals (77% of clearers), HBV DNA levels were low (<100 IU/ml) throughout the period of HBsAg clearance.

Rate of HBsAg clearance

HBsAg clearance occurred over a median time of 157 weeks (95% CI 90-239 weeks) (Fig 4A). Comparing individuals on treatment (n=4) vs. off treatment (n=9) during or in the 12 months prior to HBsAg clearance, clearance occurred over similar time frames (median 150 weeks in those on

treatment vs. 157 weeks in those not on treatment; Fig 4B). Among 279 HBsAg 'non-clearers', 246/279 (88%) had HBsAg levels that were persistently >1000 IU/ml. The remaining 12% had more heterogenous HBsAg dynamics, including transient dips <1000 IU/ml (e.g. HBS-298) and sustained levels <1000 IU/ml but without a trend towards clearance (e.g. HBS-368).

Treatment status of HBsAg clearers vs non-clearers

During the HBsAg clearance phase, or in the 12 months prior, 4/13 (31%) individuals defined as having completely cleared HBsAg were on NA therapy (Fig 4B,C). These individuals had received treatment for a median of 13 months (range 2 months – 8 years) prior to clearance. The other nine (69%) were not on treatment in the 12 months prior to HBsAg clearance, but one had received PEG-IFNα therapy 4 years earlier. In those individuals defined as 'potential clearers', 7/27 (26%) received NA treatment, two of whom received TDF as part of an HIV treatment regimen.

We also reviewed treatment data for the 279 individuals who did not clear HBsAg, and were able to retrieve data for 171 of these (61%). Among these, 131 (77%) had received treatment of some type, and 40 had never been treated (23%). We were not able to determine robust time-frames for most treatment episodes. Based on these data, non-HBsAg clearers were statistically more likely to be on treatment than HBsAg clearers (131/171, vs 4/13 respectively, p=0.001 by Fisher's Exact test). This may reflect inherently better immune control in the group who clear HBsAg, meaning they are less likely to meet criteria for treatment than non-clearers. However, these data must be interpreted with caution, as bias is introduced as a result of missing data among the non-clearers, and by different time-lines for follow-up (we assessed treatment cross-sectionally in clearers based on a specific time of HBsAg loss, for which there is no equivalent among non-clearers, thus we may have assessed longer follow-up times in the latter group).

HBeAg status

HBeAg was detectable in 81/319 (25%) of individuals at the start of the observed time period. Among these, 51/81 (63%) were male and the median age was 34. By multivariable analysis, Chinese ethnicity was associated with HBeAg-positive status, with 22/56 (39%) of Chinese

individuals being HBeAg-positive (p=0.025). We documented HBeAg clearance in 44/81 (54%) of these individuals over the observed time period (Table 3) at a median age of 37 years. HBeAg loss occurred over a median period of 54 weeks (95% CI 38-66 weeks) between last positive and first negative HBeAg test (Fig 4D). Median clearance was 49 weeks (95% CI 29-59 weeks) for individuals who had received treatment in the year prior to the last positive HBeAg result (n= 24, 55%) and 52 weeks (95% CI 14-133) for untreated individuals (n=19, 43%); treatment data were not available for 1 individual (Fig 4E,F). We also reviewed treatment data for those who did not clear HBeAg, and were able to retrieve data for 27 of these (73%). Of these, 24 (89%) had received some treatment whilst 3 (11%) were untreated.

Association between HBsAg clearance and ALT

Complete longitudinal ALT data are shown for each individual in Suppl Figs 1-3. We investigated whether there were differences in ALT according to HBsAg clearance (for each of the three HBsAg groups defined in Fig 2). There was no significant difference in ALT at the time of first test between HBsAg clearers, 'potential clearers' and non-clearers (data not shown). ALT data were available before and during HBsAg clearance for 11/13 individuals who cleared HBsAg. Among these, three individuals (HBS-162, HBS-195 and HBS-314) had a spike in ALT before clearance which returned to the normal range after HBsAg clearance. Another individual (HBS-230) also had a slightly raised ALT before HBsAg clearance, but this did not normalise after HBsAg clearance. In the 7 other cases, ALT results remained within the local reference range (10-45 iu / L) for the entire period of surveillance (Suppl Fig 1).

Relationship between HBsAg and HBV DNA

In 11/13 HBsAg clearers, HBV DNA was below the limit of detection (<20 IU/ml) throughout; in two cases, HBV DNA was cleared at the same time as HBsAg (Suppl Fig 1). The HBV DNA trajectory of individuals classified as potential clearers was more heterogenous (Suppl Fig 2): 10 individuals had cleared HBV DNA by the time of their last HBsAg test, 9 had negative HBV DNA results at some point but had subsequent detectable viraemia, and 8 individuals had detectable HBV DNA throughout the period of surveillance.

DISCUSSION

Novelty

HBsAg clearance in CHB is an uncommon event, and large cohorts over a long period of clinical follow up are therefore required to describe the characteristics of individuals who clear, and to determine the specific dynamics of serological changes. Although there have been previous studies reporting HBsAg loss, our literature review confirms that these are mostly focussed in Asia, and that relatively few studies track longitudinal data in an unbiased way. Our current approach adds novelty in a variety of ways:

i. We apply a new bespoke, algorithmic approach to collating a large longitudinal clinical dataset from multiple electronic sources. This allowed us to make use of data that are generated by routine clinical laboratories, but are not routinely used for patient care such as quantiative measures of HBsAg and HBeAg. This method of data collection also facilitated the robust identification and exclusion of duplicate patient records.

ii. Undertaking this analysis in a UK-based cohort provides a novel and more diverse mixture of host ethnicities (and by inference, diverse viral genotypes). To our knowledge this is one of the only studies of this kind in such a population.

iii. We report ≥2 HBsAg timepoints for each individual, providing long periods of clinical follow-up and the opportunity to track uncommon clearance events over time;

iv. Unlike some previous studies of HBsAg clearance that introduce bias through a focus on treatment or based on patient recall for follow-up, the approach we took is agnostic to other parameters, thereby providing a more inclusive picture of all individuals with CHB infection;

v. In addition to reporting longitudinal data for HBsAg loss, we also track HBeAg loss over time. HBeAg loss is an important immunological event (34) signifying control (typically in association with a fall of HBV DNA levels), and may also be an important target for interventions at a population level (35).

The value of the NIHR HIC approach

The NIHR HIC approach, involving the generation of standardised datasets based upon routinely-collected data, focussed on the needs of researchers in particular clinical and/or therapeutic areas, supports the re-use of tools, data, and insights across multiple research projects and organisations. As we address a wider range of questions, and as we continue to share expertise with other university-hospital partnerships, we will increase the breadth, depth, and quality of the dataset, covering a wider range of variables, for a larger patient population, and recording more information about the provenance and interpretation of the values obtained. All of the data used for this paper, and our understanding of that data, will be available for use by other researchers. Any questions that we were unable to fully address, and other questions that emerged during the analysis, will help to inform the future development of the dataset.

Role of treatment in HBsAg clearance

Our dataset corroborates prior literature in confirming that treatment is not pre-requisite for clearance, and that immunological clearance of HBsAg and HBeAg can occur independently of antiviral therapy (Fig 1C) (23, 36). There are multiple host and viral factors influencing outcome during CHB infection including host factors such as age, obesity, gender and diabetes along with genetic variations in CD8+ T-cell responses (mediated by HLA genotype), T-cell receptor antagonism and viral escape mutations (37, 38).

Due to small numbers, we did not have statistical power to determine whether there was a significant difference in the time taken to clear either HBsAg or HBeAg in individuals on treatment compared to an untreated group. However, the comparable speed of clearance on vs off treatment suggests that clearance trajectories are similar irrespective of NA treatment. We found that NA treatment was more common among non-clearers, which may reflect a genuinely higher proportion of this group meeting treatment criteria, but may also be biased by the incomplete nature of our treatment data. Further prospective studies are needed to study the relationship between clearance and treatment in more detail.

Timing of HBsAg and HBeAg clearance

Based on the epidemiology of HBV infection in this cohort, in which a substantial proportion of individuals are likely to have been infected at birth or in early childhood, it is intriguing that HBsAg and HBeAg clearance occur apparently at random in middle adulthood. In the case of HBsAg, clearance, its association with older age has been previously reported in studies we identified through our literature review (20, 23, 29). The chances of clearance may be cumulative over time; individuals infected for longer periods of time could thus have a higher chance of clearance, explaining why individuals who clear infection are, on average, older than non-clearers.

HBeAg clearance occurred over a median period of 54 weeks, substantially more quickly than HBsAg clearance which was documented over a median period of 157 weeks, perhaps indicating different underlying mechanisms at play (34, 37, 39). Further studies are needed to determine the relevant immune responses that underpin this clearance, and to identify possible triggers for clearance.

Relevance of HBsAg and HBeAg for clinical practice and research

While some guidelines recommend monitoring of HBsAg levels (1, 40), there is a lack of consistent understanding about how to interpret individual or longitudinal measurements. Developing better insights into the prognostic information that can be captured from this biomarker could be relevant to predicting patient outcomes and providing stratification of therapy. In this study, we did not have routine access to HBsAg levels >1000 IU/ml, but as these data progressively become available, future studies will have the opportunity to develop a better picture of HBsAg distribution across the whole range of CHB infections. Advocacy is required to provide more universal access to platforms that quantify HBsAg, and to improve clinical practice through interval measurements of HBsAg in chronically infected patients.

The picture we have developed here suggests that the majority of individuals who develop a sustained pattern of HBsAg decline below 1000 IU/ml are likely to go on to clear HBsAg, consistent with previous longitudinal surveillance suggesting that baseline HBsAg levels may be a more accurate prognostic marker than HBV viral load (28, 29). Prospective studies of large HBV cohorts

are likely to be needed to identify individuals on a clearance trajectory; enhanced surveillance of these individuals is a promising future route to understanding the immunological correlates of HBsAg clearance.

Caveats and limitations

Routinely-collected clinical data may be lacking in context, consistency, and completeness. Health professionals recording information to support decision-making and continuity of care, and the systems that they use, may fail to record additional, contextual information needed to address specific research questions. Variations in practice may mean that data from different sites, or from different clinicians, are incompatible (see Table 4). A collaborative approach to data quality improvement, with substantial, local clinical engagement, will help to address these challenges, but there is always more to be done, not least, as practice changes and new research questions evolve. For this paper, the questions that we were able to ask, and the size of the population considered, were limited by the nature and means of the data recorded, rather than by the basic availability. We considered only those clinical records for which the data were sufficiently complete, and for which the context was adequately explained, accepting the possibility that our exclusion of other records could introduce systematic bias.

Future questions

Prospective surveillance is important in order to provide the opportunity for studying relevant immune responses during the clearance phase. As we have shown that clearance is a relatively long process, occurring over a median of 54 weeks for HBeAg and 157 weeks for HBsAg, this provides a window of opportunity for sampling and follow-up. There is an important distinction to be made between functional cure (sustained loss of HBsAg) and sterilising cure (loss of cccDNA integrated into hepatocyte nuclei), and interest in how to determine these different outcomes. There are currently many new therapeutics in clinical trials aimed at targeting cccDNA directly including capsid effectors, RNA interference and gene editing (7). Further work is needed to develop biomarkers that can detect cccDNA in order to distinguish between these two different outcomes.

Studies of both host and viral genetics are required to underpin a better understanding of the mechanisms of clearance, including new approaches to generating full length deep sequencing of HBV, and unbiased methods to study host genetic polymorphisms that impact on disease outcome. In order to power such studies sufficiently to detect relevant signals, large collaborative multi-centre studies may be required. As we improve our insights into the dynamic changes of serological markers, opportunities arise for improving prognostication and providing better patient-stratified care.

MATERIALS AND METHODS

Health Informatics Collaborative Infrastructure

The UK National Institute for Health Research (NIHR) Health Informatics Collaborative (HIC) (www.hic.nihr.ac.uk) is a programme of infrastructure development aimed at increasing the quality and availability of routinely-collected clinical data for translational research. Eighteen university-hospital partnerships across England have signed a framework data sharing agreement, and are working to facilitate the sharing and re-use of data across centres, for approved research purposes. A key component of the NIHR HIC approach is the creation of standardised datasets to support research in specific therapeutic areas, with relevant variations in context and practice recorded as structured metadata, facilitating re-use at scale. Viral hepatitis was selected as one of the initial areas for infrastructure development: in Oxford, this has led to the establishment of new data flows from clinical and laboratory systems, the design of new screens for data capture, and the implementation of several, programmatic (and re-usable) data transformations.

Clinical cohorts

Our HBV cohort was collected from the records of a large UK teaching hospital in Oxford (http://www.ouh.nhs.uk/), which provides 1 million patient contacts per year and receives laboratory samples from the community and four inpatient sites. We retrospectively identified individuals aged ≥18 at time of database interrogation (26-Mar-2018) with chronic HBV infection (defined as positive HBsAg on ≥2 occasions ≥6 months apart) based on laboratory data collected

between January 2011 and December 2016. Inclusion criteria and other case definitions are set out in Table 1. It is standard practice in Oxfordshire to start patients on HBV treatment based on UK NICE guidelines which determine treatment eligibility using age, sex, ALT, HBV DNA and fibroscan score (40).

Data collection

Our cohort was initially defined by an electronic search of the microbiology laboratory information management systems (LIMS) to identify individuals with a positive HBsAg test. Individual subjects were allocated a pseudo-anonymised ID prefixed 'HBS', these ID numbers are included in the text to allow relevant results to be identified from within our metadata table (Suppl data table 2). We generated a data specification using search terms (Table 5) to define the data set. The Oxford NIHR Biomedical Research Centre (BRC) data warehouse receives data from operational systems within the hospital, such as electronic patient records (EPR) and LIMS (Fig 5). Within the warehouse, the data are linked, transformed, and reorganised to better support the generation of data products focussed upon a particular purpose or research area. In this case, the data product is a database containing de-identified information on patients with hepatitis. These data were cleaned and individuals not meeting inclusion criteria (Table 1) were removed.

We devised classification criteria for HBsAg and HBeAg to sort each individual into a category based on the dynamics of these serologic markers (Table 1). For HBsAg and HBeAg 'clearers' and HBsAg 'potential clearers', data which were not captured electronically or were not available from the data warehouse e.g. (most recent transient elastography score and HBV treatment status), were retrieved from the patient's written clinical record or from dictated letters from the viral hepatitis clinic.

Ethics

The NIHR HIC Viral Hepatitis database was approved by the NRES Committee South Central-Oxford C on 6th October 2015 (REC reference: 15/SC/0523).

Statistical analysis

We cleaned and analysed data using R and the data.table package (41). The clearance rate was calculated as $\frac{Number\ of\ patients\ who\ cleared}{Total\ patient\ years} \times 100$. Plots were created using ggplot2 (42), and survival analysis and Kaplan-Meier plots created using the survival and rms packages (43). We used Wilcoxon or Kruskal Wallis tests for mean comparison of continuous variables, Fisher's exact test for comparison of categorical variables, and logistic regression for multivariable analysis. We included all the parameters in our dataset in multivariable analysis, based on existing biological reasons to believe them likely to be relevant. Specifically, age at first HBsAg test is known to be associated with HBsAg clearance, sex and ethnicity could indicate diffrences in host genetics and immune response, and HBeAg status is a known marker of disease severity (20, 26, 37). Code used for this analysis is available in the attached HBsAg_Final_Analysis.html file included in the supplementary information. To define HBsAg clearance time-frames, we measured from the time of the last HBsAg result of >1000 IU/ml (or the result closest to 1000 IU/ml) to the time point at which HBsAg first became undetectable. For analysis of ALT, we used the result corresponding to the time of the first HBsAg test result.

TABLES

Table 1: Summary of criteria used to confirm inclusion in the analysis and to classify individuals according to HBsAg dynamics and HBeAg dynamics

Category	Criteria
Inclusion in cohort for analysis	• Unique electronic record available • Age ≥18 at time of data interrogation • Longitudinal laboratory data available • No ambiguous data points[a] • HBsAg detectable at ≥2 timepoints ≥ 6 months apart (HBsAg >20 IU/ml) • ≥1 futher HBsAg reading (either positive or negative) with a total surveillance period of ≥12 months
HBsAg categories	
HBsAg clearer	• HBsAg initially detectable, but subsequently falls below the limit of detection (<20 IU/ml) • HBsAg does not rebound to ≥20 IU/ml • ≥2 consecutive HBsAg readings <20 IU/ml
Potential HBsAg clearer	• HBsAg falls <1000 IU/ml on ≥2 independent occasions, • HBsAg does not rebound to >1000 IU/ml • HBsAg not below the limit of detection for two consecutive readings
Non HBsAg clearer	• All individuals who are not classified as HBsAg clearer or potential clearer
HBeAg categories	
HBeAg persistently positive	• HBeAg above the limit of detection (≥20 IU/ml) for all timepoints.
HBeAg persistently negative	• HBeAg below the limit of detection (<20 IU/ml) for all timepoints.
HBeAg clearer	• HBeAg detectable at ≥2 independent timepoints and subsequently falls below the limit of detection for ≥2 consecutive timepoints • HBeAg does not rebound above the limit of detection
Non HBeAg Clearer	• All individuals who are not classified as persistently HBeAg positive, negative or as an HBeAg clearer

[a] Records with free text or uninterpretable data were removed from analysis

Table 2: Baseline characteristics of 319 individuals with chronic HBV infection recruited through a UK cohort and classified according to pattern of HBsAg clearance over time.

	Whole cohort	HBsAg clearers and potential clearers	HBsAg non-clearers	p-value (uni-variable analysis)	p-value (multi-variable analysis)	Effect Size	Std Error
Number of individuals	319	40	279	NA	NA		
Median age in years at time of first HBsAg test	34	40	34	0.0034*	0.0081*	-0.4	0.01
Sex (%)				0.605	0.605	0.14	0.36
Male	191 (60)	26 (65)	165 (59)				
Female	128 (40)	14 (35)	114 (41)				
Self-reported ethnicity (%)							
White	92 (29)	15 (38)	77 (28)	0.649	0.632	-0.55	0.55
Mixed	18 (6)	0 (0)	18 (6)	0.991	0.989	15.39	898.69
Asian or Asian British	52 (16)	7(18)	45 (16)	0.641	0.705	-0.25	0.63
Black or Black British	46 (14)	5 (12)	41 (15)	0.697	0.638	-0.06	0.68
Chinese	56 (18)	8 (20)	48 (17)	0.902	0.914	-0.28	0.62
Any Other Ethnic Group	7 (2)	0 (0)	7 (3)	0.992	0.994	15.39	1458.87
Not Stated	48 (15)	5 (12)	43 (15)	NA	NA	NA	NA
HBeAg positive status at baseline (%)	81 (25)	6 (15)	65 (23)	0.3105	0.131	0.51	0.49
Median elastography score, kPa (most recent value)	5.3	4.5	5.5[a]	0.18	NA	NA	NA
Number of patients receiving treatment (%)	142/211[b] (67)	11/40[c] (28)	131/171[b] (76)	NA	NA	NA	NA

[a] Elastography data available for 42 individuals in the non-clearance group, as data not routinely recorded electronically.

[b] Treatment data were missing for 108 individuals among the HBsAg non-clearers, as data not routinely recorded electronically.

[c] Treatment in the 12 months before the last positive HBsAg test

NA = not applicable

Table 3: Baseline characteristics of HBeAg positive individuals classified according to HBeAg clearance over the observed time period

	HBeAg clearers	HBeAg non-clearers	p-value (uni-variable analysis)
Number of individuals	44	37	
Median age in years at time of first HBsAg test	34	35	0.75
Sex (%) Male Female	 29 (66) 15 (34)	 25 (58) 12 (32)	1
Self-reported Ethnicity (%) White Mixed Asian or Asian British Black or Black British Chinese Any Other Ethnic Group Not Stated	 12 (27) 4 (9) 8 (18) 6 (14) 8 (18) 1 (2) 5 (11)	 10 (27) 4 (11) 3 (8) 0 (0) 14 (38) 2 (5) 4 (11)	 0.959 0.819 0.427 0.991 0.330 0.512 NA
Median elastography score, kPa (based on most recent value)	5.5	4.55	0.24
Number of patients receiving treatment [a] (%)	24/44 (55)	24/27[b] (89)	NA

NA = not applicable

[a] Treatment in the 12 months prior to the last positive HBeAg result

[b] Treatment data were missing for 10 individuals among the HBeAg non-clearers as data not routinely collected electronically.

Table 4: Factors influencing the analysis of retrospective clinical HBV data

Category of influence	Examples of the effect on data integrity
Patient factors	• Many individuals with CHB infection globally are not diagnosed; those with data available for clinical analysis represent a distinct minority group who have been able to access healthcare and follow-up (44). • Patients are lost to follow-up or move between regions. • HBV diagnosis rarely occurs in acute infection, so the duration of infection prior to clearance is unknown. • HBsAg clearance is a relatively infrequent event and thus patient numbers for analysis are small. • Description of a changing cohort is challenging e.g. age changes over time, patients start and stop therapy.
Healthcare factors	• Different assays are not always requested simultaneously, thus limiting the correlation between variables (e.g. HBV DNA vs HBsAg). • Follow-up occurs over a variety of different time frames, with different intervals between follow up visits; clearance durations may therefore be over-estimated due to infrequent sampling. • Treatment can alter the dynamics of biomarkers (e.g. ALT, HBV DNA).
Laboratory factors	• Assay platforms change over time, which may alter sensitivity, specificity and limits of detection. • Quantitative assays have upper and lower limits of quantification; values outside the window of detection cannot be analysed. • False positive or false negative tests may occur. • Certain data are not routinely generated or captured (e.g. HBV genotype).
Data factors	• Results are captured by a variety of different electronic systems (electronic patient record, electronic laboratory systems, pharmacy systems, hand-written clinical notes, dictated clinic letters). • Different healthcare professionals may not record data consistently and coding is subject to errors. • Free text entries in laboratory reporting can lead to errors or ambiguities (e.g. use of comma vs. full stop for decimal point). Certain parameters are not consistently recorded, e.g. ethnicity. • The electronic pipeline only collects certain pre-defined data (e.g. for HIV, HCV, HDV we were only able to access viral load data, not antibody tests, and therefore we do not know the denominator of total tests performed). • Treatment data may not be recorded electronically (often recorded as part of paper notes, making them more difficult to trace); start dates often not documented for patients on long-term treatment. • Poor continuity of data when patients are transferred between different healthcare providers.

Table 5: 'Data dictionary' of clinical and demographic parameters collected for cohort of individuals with chronic HBV infection

Laboratory Parameter	Data source	Date range (for laboratory parameters)	Assay platform	Notes
HBsAg	Microbiology LIMS (Sunquest)	09/2004 – 03/2018	Centaur; 09/2004 – 12/2014 Abbott Architect i2000SR (Abbott laboratories, Chicago, IL); 12/2014 – 03/2018	Traditionally reported as binary test (positive/negative) but generates semi-quantitative data. Lower limit of detection 0.05 IU/mL.
HBeAg	Microbiology LIMS (Sunquest)	04/1995 - 03/2018	Centaur; 09/2004 – 12/2014 Abbott Architect i2000SR (Abbott laboratories, Chicago, IL) 12/2014 – 03/2018	Traditionally reported as binary test (positive/negative) but generates semi-quantitative data.
HBV DNA	Microbiology LIMS (Sunquest)	03/2009 – 03/2018	Cobas TaqMan assay (Roche diagnostics, Branchburg, NJ)	Lower limit of detection 0.9x10^1IU/mL. 1 IU/ml is equivalent to 2.5-5 genome equivalents (copies/ml)
ALT	Biochemistry LIMS (LIMS)	02/2013 – 01/2018	Siemens ADVIA 2400; 02/2013 – 01/2015 Abbott Architect c16000 or c8000 (Abbott laboratories, Chicago, IL); 01/2015 – 01/2018	Reported as quantitative value. Normal reference range 10-45 IU/L*
Ethnicity	Hospital EPR (Cerner Millennium)	NA	NA	Self-reported according to standardised ethnicity codes
Fibroscan result (transient elastography score)	Hospital EPR (Cerner Millennium) / Clinic letter database (Manual)	NA	EchoSens, Paris	Most recent recorded elastography result
HBV treatment status	Hospital EPR (Cerner Millennium) / Clinic letter database (Manual)	NA	NA	Treatment guidelines changed over time, so use of different agents applied across the timespan of the cohort.

LIMS = Laboratory information management system; EPR = Electronic patient record

* In our hospital, no distinction is made in the ALT reference range for males vs females

FIGURE LEGENDS:

Fig 1: Cartoons depicting key pathways in HBV replication cycle to illustrate targets that may bring about control or clearance.

A: Pathways relevant to maintenance of HBV infection. HBV viral DNA is released in the nucleus, and cccDNA is formed by covalent ligation of the two DNA strands. A stable mini-chromosome is formed, allowing persistence of the virus over time. The cccDNA acts as the template for mRNA and pregenomic RNA (pgRNA). Viral reverse transcriptase (RT) generates new genomic DNA from pgRNA. Non-infectious sub-viral particules (SVP) form from HBsAg and new infectious virions assemble, for release into the blood stream. HBsAg measurement accounts for both the SVP and infectious virions, whereas infectious virions alone can be measured through HBV viral load (HBV DNA).

B: Pathways relevant to suppression of HBV infection by NA therapy: Inhibiton of viral RT suppresses generation of new viral DNA. This means new infectious HBV virions cannot be constructed and HBV DNA is undetectable in plasma. However, cccDNA remains as a persistent reservoir in the hepatocyte nucleus, so HBsAg production can continue and rebound viraemia is likely following cessation of therapy. For this reason, individuals with CHB on successful treatment frequently have an undetectable viral load but remain HBsAg-positive.

C: Pathways relevant to functional or sterilising cure of HBV infection: Upregulation of host immune responses or therapy with interferon (IFN) leads to elimination of the persistent cccDNA reservoir either through death of the hepatocyte or unknown non-lytic methods. HBsAg and HBV DNA both disappear from the blood stream. In practice, there is no clinical test that can confirm complete ('sterilising') cure, so this group is usually regarded as being at a small risk of relapse (i.e. 'functional' cure).

Fig 2: Flowchart showing identification and classification of adults with chronic HBV infection from a hospital electronic system. The figure represents 319 individuals who met

inclusion criteria, and divides these into three different categories according to HBsAg clearance, and four categories for HBeAg; (for classification criteria, see Table 1).

Fig 3: Exemplar trajectories of HBsAg over time representing adults with chronic HBV infection. Individuals are classified as (A) a complete HBsAg clearer, (B) a potential HBsAg clearer (C) a non-HBsAg clearer; (for classification criteria, see Table 1).

Fig 4: Kaplan-Meier curves showing trajectory of HBsAg clearance (N=13) and HBeAg clearance (N=43) for selected individuals who met criteria for complete clearance from within a cohort of adults with chronic HBV infection. Data are shown for HBsAg (panels A-C) and for HBeAg (panels D-F), initially for all clearers (panels A and D), and then subdivided according to treatment status (panels B and E). Boxes C and F report the median time to clearance for each group in weeks, with 95% confidence intervals. For HBsAg clearance, the upper confidence interval for treated cases cannot be determined due to small numbers. Treatment of HBsAg clearers and potential clearers comprised TDF monotherapy (n=3), TDF with emtricitabine (n=2), 3TC with ADV or TDF (n=4), 3TC monotherapy (n=1), ETV montherapy (n=1). Treatment of HBeAg clearers comprised TDF monotherapy (n=10). 3TC monotherapy (n=2) ETV monotherapy (n=5), 3TC with ADV (n=3), IFN with RBV (n=1), IFN monotherapy (n=3), treatment data were not available for one individual. * When no values >1000 IU/ml were recorded, the highest value was used. ** Not enough data to cacluate upper CI. § Treatment status not known for 1 individual.

Fig 5: Flow diagram to depict collection, storage and output of electronic clinical data from a Health Informatics Collaborative data warehouse.

The data warehouse receives data from operational systems within the hospital such as electronic patient records and laboratory information management systems (LIMS) and maps this data to individuals where the identifiers are then stored in the master data store and provides the mappings for data products. De-identified linked data is stored separately and forms the content of data products. Definitions of data items are recorded in the metadata catalogue. Data items for the

data product are selected using the definitions in the metadata catalogue the mappings for these are retrieved from the master data store and data retrieved from the integrated data store to create the final data product.

SUPPLEMENTARY DATA:

On acceptance for publication, Supplementary Data will be made available at DOI: 10.6084/m9.figshare.7262957. Prior to publication, these files can be accessed using the following URL: https://figshare.com/s/82db3b5cd1dc5c6dd566

Supplementary Table 1: **Summary of studies reporting rate and quantitation of HBsAg clearance in individuals with HBV infection -** Listed studies were identified from a PubMed search performed in April 2018 with the search terms: ('hepatitis B' OR 'HBV') AND ('clearance' OR 'seroclearance' OR 'vir* negative'), written in English between 2008 – 2018, and reporting a cumulative or annual clearance rate of HBsAg using a quantitative assay.

Supplementary Table 2**: Data for 319 adults with chronic HBV infection**

Supplementary Figure 1: **Longitudinal data for 13 adults with chronic HBV infection who completely cleared HBsAg -** Each individual is labelled with a unique anonymised ID number, prefixed HBS. Time is shown in weeks since the first HBsAg positive test. Units (y-axis) are shown in IU/ml, except for ALT which is shown in IU/L.

Supplementary Figure 2: **Longitudinal data for 27 adults with chronic HBV infection on a potential HBsAg clearance trajectory.** Each individual is labelled with a unique anonymised ID number, prefixed HBS. Time is shown in weeks since the first HBsAg positive test. Units (y-axis) are shown in IU/ml, except for ALT which is shown in IU/L.

Supplementary Figure 3: **Longitudinal data for 279 adults with chronic HBV infection who did not clear HBsAg**. Each individual is labelled with a unique anonymised ID number, prefixed HBS. Time is shown in weeks since the first HBsAg positive test. Units (y-axis) are shown in IU/ml, except for ALT which is shown in IU/L.

Supplementary Figure 4: **Boxplot showing the distribution of age among individuals who clear or potentially clear HBsAg (n=40, median age 40) and those who do not clear HBsAg (n=279; median age 34).**

Supplementary Code: html file

REFERENCES

1. EASL. 2017. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. J Hepatol 67:in press.
2. McNaughton AL, DArienzo V, Azim Ansari M, Lumley SF, Littlejohn M, Revill P, McKeating JA, Matthews PC. 2019. Insights From Deep Sequencing of the HBV Genome-Unique, Tiny, and Misunderstood.
3. Schinazi RF, Asselah T. 2017. From HCV To HBV Cure. Liver Int 37:73–80.
4. Lok AS, Zoulim F, Dusheiko G, Ghany MG. 2017. Hepatitis B cure: From discovery to regulatory approval. Hepatology 66:1296–1313.
5. Perrillo RP, Martin P, Lok AS. 2015. Preventing hepatitis B reactivation due to immunosuppressive drug treatments. JAMA - J Am Med Assoc 313:1617–1618.
6. Perrillo RP, Gish R, Falck-Ytter YT. 2015. American Gastroenterological Association Institute Technical Review on Prevention and Treatment of Hepatitis B Virus Reactivation During Immunosuppressive Drug Therapy.
7. Lucifora J, Xia Y, Reisinger F, Zhang K, Stadler D, Cheng X, Sprinzl MF, Koppensteiner H, Makowska Z, Volz T, Remouchamps C, Chou W-M, Thasler WE, Hüser N, Durantel D, Liang TJ, Münk C, Heim MH, Browning JL, Dejardin E, Dandri M, Schindler M, Heikenwalder M, Protzer U. 2014. Specific and Nonhepatotoxic Degradation of Nuclear Hepatitis B Virus cccDNA. Science (80-) 343:1221 LP-1228.
8. Jaroszewicz J, Serrano BC, Wursthorn K, Deterding K, Schlue J, Raupach R, Flisiak R, Bock CT, Manns MP, Wedemeyer H, Cornberg M. 2010. Hepatitis B surface antigen (HBsAg) levels in the natural history of hepatitis B virus (HBV)-infection: A European perspective. J Hepatol 52:514–522.
9. WHO. 2017. Hepatitis B. Fact Sheet.
10. Seto W-K, Wong DK-H, Fung J, Ip PPC, Yuen JC-H, Hung IF-N, Lai C-L, Yuen M-F. 2012. High hepatitis B surface antigen levels predict insignificant fibrosis in hepatitis

B e antigen positive chronic hepatitis B. PLoS One 56:e43087.

11. Xun Y, Zang G, Guo J, Yu X, Liu H, Xiang J, Liu J, Shi J. 2013. Serum hepatitis B surface antigen quantification as a useful assessment for significant fibrosis in hepatitis B e antigen-positive hepatitis B virus carriers. J Gastroenterol Hepatol 28:1746–1755.

12. Martinot-Peignoux M, Carvalho-Filho R, Ferreira AC, Cardoso N, Lapalus M, Lada O, Asselah T, Krause F, Marcellin P, Murray KF, Szenborn L, Wysocki J, Rossi S, Corsa AC, Dinh P, Mchutchison JG, Pang PS, Luminos LM, Pawlowska M, Mizerski J. 2012. Significant genotype-specific association of Hepatitis B surface antigen level and severity of liver disease in patients with chronic Hepatitis B. J Hepatol 56:S211.

13. Chen CJ, Iloeje UH, Yang HI. 2007. Long-Term Outcomes in Hepatitis B: The REVEAL-HBV Study. Clin Liver Dis 11:797–816.

14. Brunetto MR, Oliveri F, Colombatto P, Moriconi F, Ciccorossi P, Coco B, Romagnoli V, Cherubini B, Moscato G, Maina AM, Cavallone D, Bonino F. 2010. Hepatitis B surface antigen serum levels help to distinguish active from inactive hepatitis B virus genotype D carriers. Gastroenterology 139:483–490.

15. Liu J, Yang H-I, Lee M-H, Jen C-L, Batrla-Utermann R, Lu S-N, Wang L-Y, You S-L, Chen C-J. 2016. Serum Levels of Hepatitis B Surface Antigen and DNA Can Predict Inactive Carriers With Low Risk of Disease Progression. Hepatology 64:381–389.

16. Huo T, Wu J-C, Hwang S-J, Lai C-R, Lee P-C, Tsay S-H, Chang F-Y, Lee S-D. 2000. Factors predictive of liver cirrhosis in patients with chronic hepatitis B: a multivariate analysis in a longitudinal study. Eur J Gastroenterol Hepatol 12.

17. Iloeje UH, Yang H, Su J, Jen C, You S, Chen C. 2017. Predicting Cirrhosis Risk Based on the Level of Circulating Hepatitis B Viral Load. Gastroenterology 130:678–686.

18. R de F, Meucci G, Vecchi M, al et. 1993. THe natural history of asymptomatic

hepatitis b surface antigen carriers. Ann Intern Med 118:191–194.

19. Manno M, Cammà C, Schepis F, Bassi F, Gelmini R, Giannini F, Miselli F, Grottola A, Ferretti I, Vecchi C, De Palma M, Villa E. 2017. Natural history of chronic HBV carriers in northern Italy: Morbidity and mortality after 30 years. Gastroenterology 127:756–763.

20. Zhou K, Contag C, Whitaker E, Terrault N. 2019. Spontaneous loss of surface antigen among adults living with chronic hepatitis B virus infection: a systematic review and pooled meta-analyses. lancet Gastroenterol Hepatol 4:227–238.

21. Hara T, Suzuki F, Kawamura Y, Sezaki H, Hosaka T, Akuta N, Kobayashi M, Suzuki Y, Saitoh S, Arase Y, Ikeda K, Kobayashi M, Watahiki S, Mineta R, Kumada H. 2014. Long-term entecavir therapy results in falls in serum hepatitis B surface antigen levels and seroclearance in nucleos(t)ide-naive chronic hepatitis B patients. J Viral Hepat 21:802–808.

22. Kim G-A, Lim Y-S, An J, Lee D, Shim JH, Kim KM, Lee HC, Chung Y-H, Lee YS, Suh DJ. 2014. HBsAg seroclearance after nucleoside analogue therapy in patients with chronic hepatitis B: clinical outcomes and durability. Gut 63:1325–1332.

23. Kobayashi M, Hosaka T, Suzuki F, Akuta N, Sezaki H, Suzuki Y, Kawamura Y, Kobayashi M, Saitoh S, Arase Y, Ikeda K, Miyakawa Y, Kumada H. 2014. Seroclearance rate of hepatitis B surface antigen in 2,112 patients with chronic hepatitis in Japan during long-term follow-up. J Gastroenterol 49:538–546.

24. Kuo Y-H, Chang K-C, Wang J-H, Tsai P-S, Hung S-F, Hung C-H, Lu S-N. 2015. Changing serum levels of quantitative hepatitis B surface antigen and hepatitis B virus DNA in hepatitis B virus surface antigen carriers: a follow-up study of an elderly cohort. Kaohsiung J Med Sci 31:102–107.

25. Nagaoka S, Abiru S, Komori A, Sasaki R, Bekki S, Hashimoto S, Saeki A, Yamasaki K, Migita K, Nakamura M, Ezaki H, Yatsuhashi H. 2016. Hepatic flares promote

rapid decline of serum hepatitis B surface antigen (HBsAg) in patients with HBsAg seroclearance: A long-term follow-up study. Hepatol Res 46:E89-99.

26. Park YM, Lee SG. 2016. Clinical features of HBsAg seroclearance in hepatitis B virus carriers in South Korea: A retrospective longitudinal study. World J Gastroenterol 22:9836–9843.

27. Jeng W-J, Chen Y-C, Chien R-N, Sheen I-S, Liaw Y-F. 2017. Incidence and predictors of HBsAg seroclearance after cessation of nucleos(t)ide analogue therapy in HBeAg negative chronic hepatitis B. Hepatology.

28. Ungtrakul T, Sriprayoon T, Kusuman P, Chunnuan P, Soonklang K, Sornsamdang G, Auewarakul CU, Tanwandee T. Role of quantitative hepatitis B surface antigen in predicting inactive carriers and HBsAg seroclearance in HBeAg-negative chronic hepatitis B patients.

29. Lim TH, Gane E, Moyes C, Borman B, Cunningham C. 2016. HBsAg loss in a New Zealand community study with 28-year follow-up: rates, predictors and long-term outcomes. Hepatol Int 10:829–837.

30. Sonneveld MJ, Hansen BE, Piratvisuth T, Jia J-D, Zeuzem S, Gane E, Liaw Y-F, Xie Q, Heathcote EJ, Chan HL-Y, Janssen HLA. 2013. Response-guided peginterferon therapy in hepatitis B e antigen-positive chronic hepatitis B using serum hepatitis B surface antigen levels. Hepatology 58:872–880.

31. Moucari R, Mackiewicz V, Lada O, Ripault MP, Castelnau C, Martinot-Peignoux M, Dauvergne A, Asselah T, Boyer N, Bedossa P, Valla D, Vidaud M, Nicolas-Chanoine MH, Marcellin P. 2009. Early serum HBsAg drop: A strong predictor of sustained virological response to pegylated interferon alfa-2a in HBeAg-negative patients. Hepatology 49:1151–1157.

32. National institute for Health and Care Excellence. 2013. Hepatitis B (chronic): diagnosis and management. Clin Guidel [CG165].

33. Griggs D, Stafford-Smith M, Gaffney O, Rockstrom J, Ohman MC, Shyamsundar P, Steffen W, Glaser G, Kanie N, Noble I. 2013. Policy: Sustainable development goals for people and planet. Nature 495:305–307.

34. Matthews PC, Carlson JM, Beloukas A, Malik A, Jooste P, Ogwu A, Shapiro R, Riddell L, Chen F, Luzzi G, Jesuthasan G, Jeffery K, Jojic N, Ndung'U T, Carrington M, Goulder PJR, Geretti AM, Klenerman P. 2016. HLA-A is a Predictor of Hepatitis B e Antigen Status in HIV-Positive African Adults. J Infect Dis 213.

35. McNaughton AL, Lourenço J, Hattingh L, Adland E, Daniels S, Van Zyl A, Akiror CS, Wareing S, Jeffery K, Azim Ansari M, Klenerman P, R Goulder PJ, Gupta S, Jooste P, Matthews PC. HBV vaccination and PMTCT as elimination tools in the presence of HIV: insights from a clinical cohort and dynamic model.

36. Habersetzer F, Moenne-Loccoz R, Meyer N, Schvoerer E, Simo-Noumbissie P, Dritsas S, Baumert TF, Doffoel M. 2015. Loss of hepatitis B surface antigen in a real-life clinical cohort of patients with chronic hepatitis B virus infection. Liver Int 35:130–139.

37. Lumley S, McNaughton A, Klenerman P, Lythgoe K, Matthews P. 2018. Hepatitis B virus Adaptation to the CD8 + T Cell Response□: Consequences for Host and Pathogen 9:1–14.

38. Matsuura K, Isogawa M, Tanaka Y. 2016. Host genetic variants influencing the clinical course of Hepatitis B virus infection. J Med Virol 88:371–379.

39. Lumley S, Noble H, Hadley MJ, Callow L, Malik A, Chua YY, Duffey OJ, Grolmusova N, Kumar A, Ravenscroft S, Spencer JI, Neumann-Haefelin C, Thimme R, Andersson M, Klenerman P, Barnes E, Matthews PC. 2016. Hepitopes: A live interactive database of HLA class I epitopes in hepatitis B virus. Wellcome Open Res 1:9.

40. 2013. Nice Clinical Guideline, Hepatitis B (chronic): diagnosis and management.

NICE Clinical Guideline.

41. Dowle M, Srinivasan A. 2018. data.table: Extension of `data.frame`. R package version 1.11.4. https://CRAN.R-project.org/package=data.table.

42. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

43. Therneau T. 2015. A Package for Survival Analysis in S_. version 2.38, <URL: Https://CRAN.R-project.org/package=survival>.

44. O'Hara GA, McNaughton AL, Maponga T, Jooste P, Ocama P, Chilengi R, Mokaya J, Liyayi MI, Wachira T, Gikungi DM, Burbridge L, O'Donnell D, Akiror CS, Sloan D, Torimiro J, Yindom LM, Walton R, Andersson M, Marsh K, Newton R, Matthews PC. 2017. Hepatitis B virus infection as a neglected tropical disease. PLoS Negl Trop Dis 11.

781

Fig 1

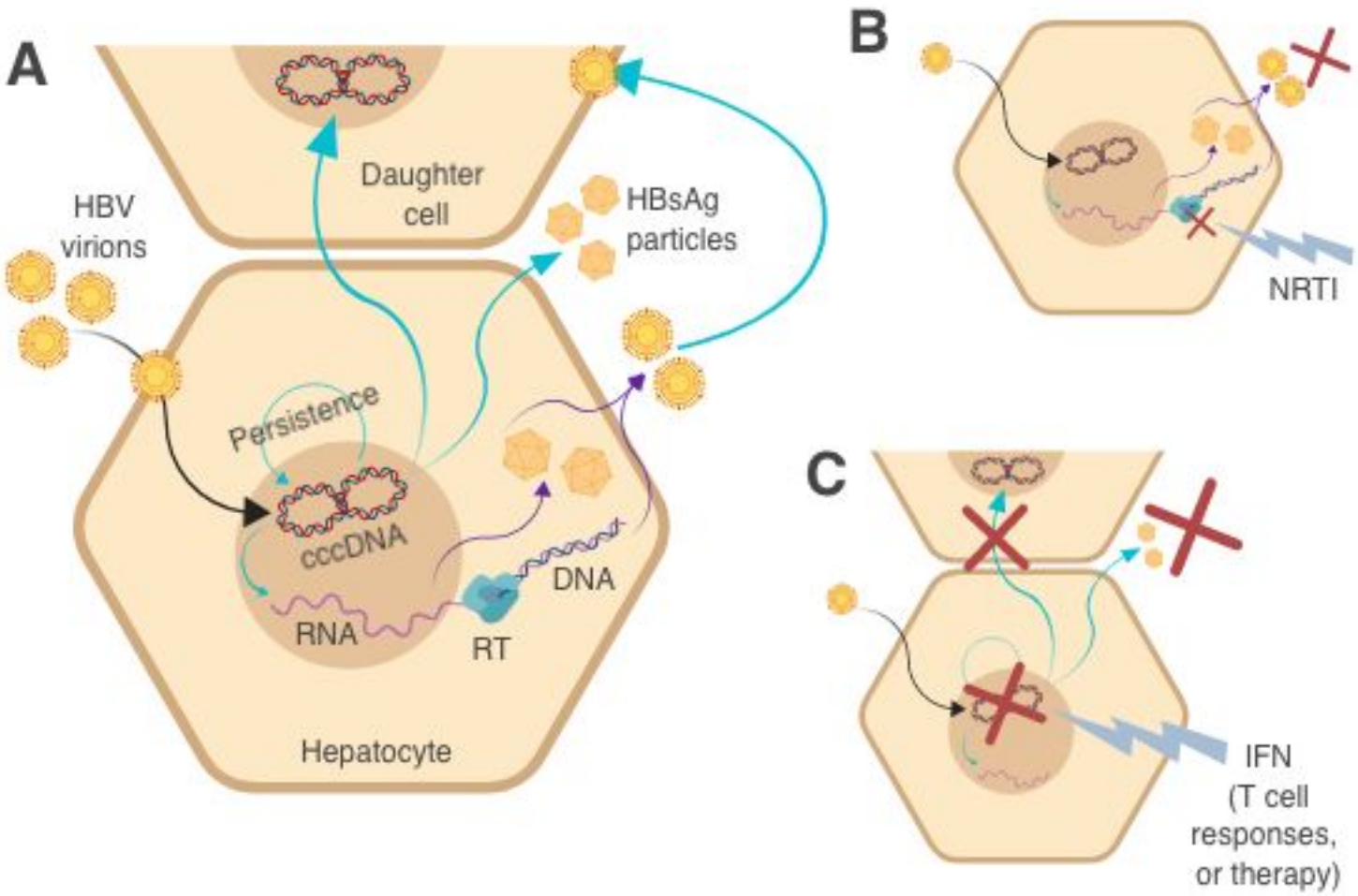

A
Daughter cell
HBV virions
HBsAg particles
Persistence
cccDNA
RNA
RT
DNA
Hepatocyte
B
NRTI
C
IFN (T cell responses, or therapy)

Fig 2

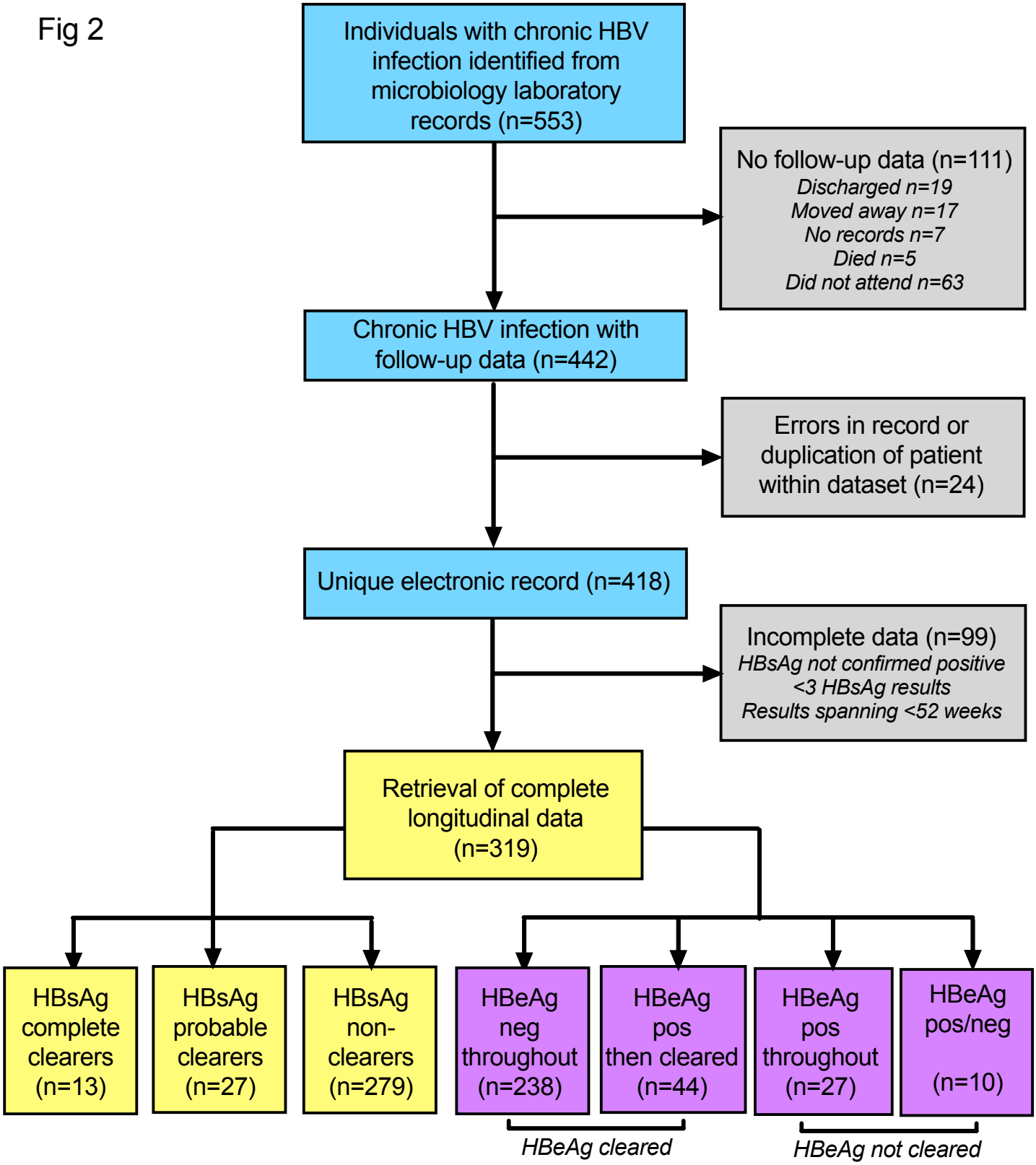

Individuals with chronic HBV infection identified from microbiology laboratory records (n=553)
No follow-up data (n=111)
Discharged n=19
Moved away n=17
No records n=7
Died n=5
Did not attend n=63
Chronic HBV infection with follow-up data (n=442)
Errors in record or duplication of patient within dataset (n=24)
Unique electronic record (n=418)
Incomplete data (n=99)
HBsAg not confirmed positive
<3 HBsAg results
Results spanning <52 weeks
Retrieval of complete longitudinal data (n=319)
HBsAg complete clearers (n=13)
HBsAg probable clearers (n=27)
HBsAg non-clearers (n=279)
HBeAg neg throughout (n=238)
HBeAg pos then cleared (n=44)
HBeAg pos throughout (n=27)
HBeAg pos/neg (n=10)
HBeAg cleared
HBeAg not cleared

Fig 3

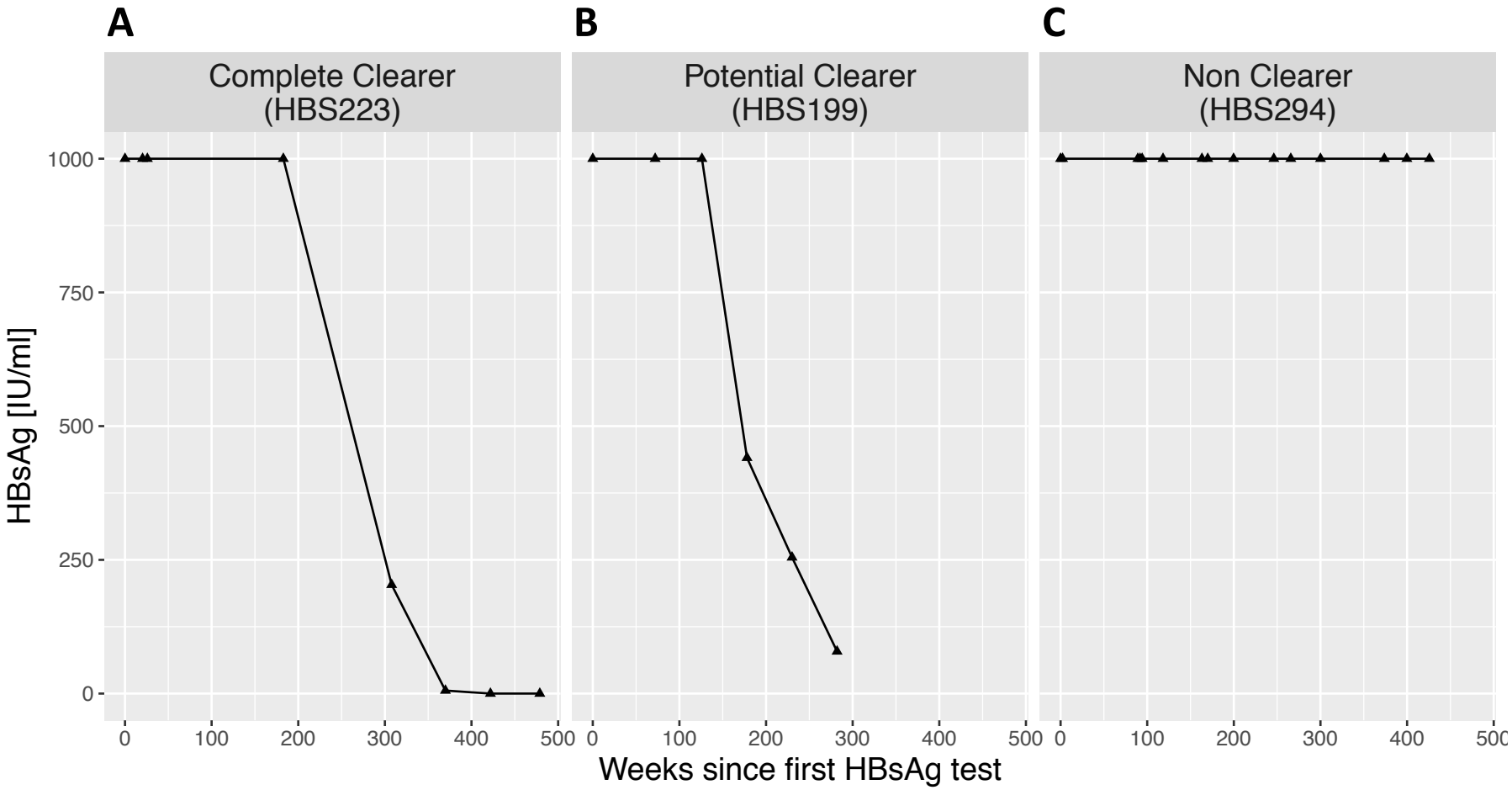
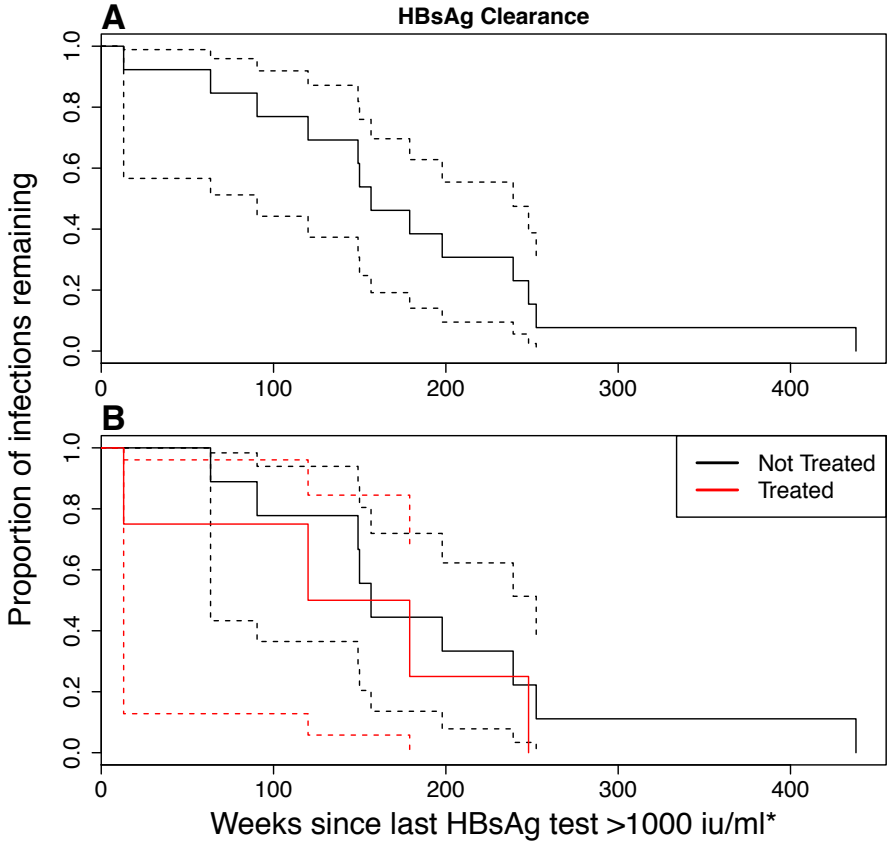
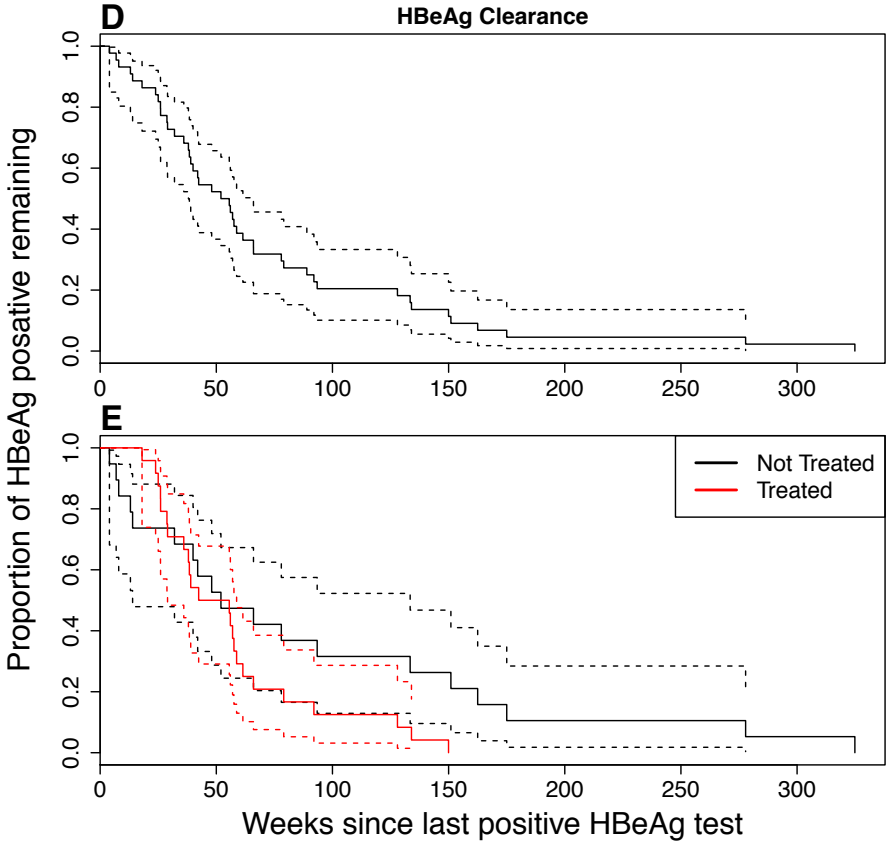

A
B
C
Complete Clearer
(HBS223)
Potential Clearer
(HBS199)
Non Clearer
(HBS294)
HBsAg [IU/ml]
1000
750
500
250
0
0
100
200
300
400
500
Weeks since first HBsAg test

Fig 4





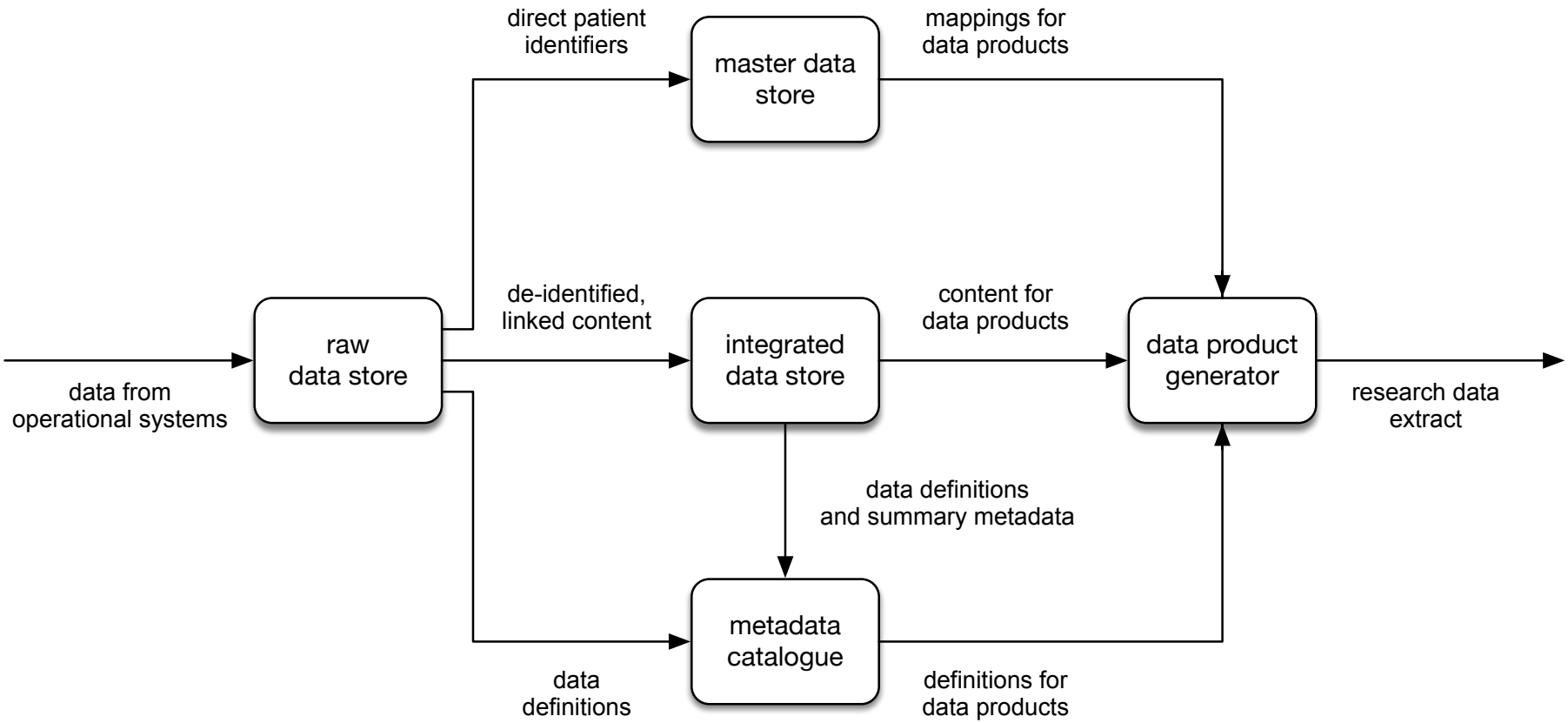
C

Time (weeks)	All cases of HBsAg clearance (n=13)	Untreated cases (n=9)	Treated cases (n=4)
Median	157	157	150
95% Confidence Intervals	90 – 239	63 - 252	13 – NA**

F

Time (weeks)	All cases of HBeAg clearance (n=43)§	Untreated cases (n=19)	Treated cases (n=24)
Median	54	52	49
95% Confidence Intervals	38 – 66	14 – 133	29 - 59

Fig 5


direct patient identifiers
master data store
mappings for data products
de-identified, linked content
content for data products
raw data store
integrated data store
data product generator
data from operational systems
research data extract
data definitions and summary metadata
metadata catalogue
data definitions
definitions for data products