Witness Name: Janet Valentine Statement No.: WITN6886001 Exhibits: WITN6886002 - 008 Dated: 21/09/2021

INFECTED BLOOD INQUIRY

WRITTEN STATEMENT OF JANET VALENTINE ON BEHALF OF THE CLINICAL PRACTICE RESEARCH DATALINK

I provide this statement on behalf of the Clinical Practice Research Datalink in response to the request under Rule 9 of the Inquiry Rules 2006 dated 2 July 2021.

I, Janet Valentine, will say as follows: -

<u>Contents</u>

Contents	. 1
Section 1: Introduction	. 3
Personal Details	. 3
Role and Responsibilities	. 3
CPRD's Data Holdings	. 3
CPRD's remit	. 3
CPRD data collection	. 4
CPRD databases and Read codes	. 5
Limitations of CPRD data	. 6
Size of the CPRD database	. 6
Changing population coverage of data over time	. 6
Adoption of EHR by GP practices	. 7
Information captured in the EHR	. 7
Collection of Read Codes	. 8
Data used to respond to the Inquiry	. 8

Section 2: Data	11
Number of people directly infected with HBV, HCV and/or HIV through the use of blood transfusions and blood products	11
The number of recipients of blood products	13
Survivorship of any individuals infected with HBV, HCV, and/or HIV through blood transfusions or blood products	16
Clinical records with a Read code that relates to blood transfusions and/or blood product by year of date associated with that event ('event date') in respect of the number of recipients of blood transfusions and blood products	:s 16
Clinical records which have at least one entry responsive to infection	18
Estimates of the prevalence or scale of infection in relation to HBV, HCV, and/or HIV	19
Estimates of prevalence or scale of infection in relation to HBV, HCV or HIV	19
Estimates of transmission rates in relation to HBV, HCV and HIV respectively	20
Search Strategy Criteria	21
APPENDICES FOR SEARCHES CONDUCTED IN CPRD GOLD AND CPRD AURUM	23
Appendix 1A: Search terms used to identify transfusions, blood products and HBV, HCV or HIV infection in the CPRD GOLD database	23
Appendix 1B: Search terms used to identify transfusions, blood products and HBV, HCV or HIV infection in the CPRD Aurum database	23

Section 1: Introduction

1. Save where otherwise stated, all facts and matters referred to in this witness statement are true and within my own knowledge or have come to my attention during the course of my work and through the research conducted by my colleagues in order to respond to the Rule 9 Request. I have carefully reviewed the analysis and I am comfortable that it is correct. Insofar as facts and matters are not directly within my knowledge, they are true to the best of my knowledge and belief and I have exhibited relevant underlying documents and data to assist the Inquiry. I am duly authorised to make this statement on behalf of the CPRD.

Personal Details

2. My personal details are as follows:

Name:	Dr Janet Valentine
Address:	Clinical Practice Research Datalink
	The Medicines and Healthcare products Regulatory Agency
	10 South Colonnade,
	Canary Wharf,
	London E14 4PU

DoB: **GRO-C** 1964 Qualifications: BSc (Hons) PhD (Medicine)

Role and Responsibilities

 My current role is the Director of the Clinical Practice Research Datalink (CPRD). I am responsible for accountability for CPRD's strategy, budget and operational delivery.

CPRD's Data Holdings

CPRD's remit

- 4. It is perhaps helpful to give a brief overview of CPRD's remit. CPRD is the UK Government's dedicated health data research service, which for more than 30 years, has been providing health data and research services to academics, medicines regulators and the pharmaceutical industry to support public health research.
- 5. CPRD's remit is to exclusively support public health research. Permission to supply data for research purposes is granted annually by the Health Research Authority. CPRD does not have permission to support non-research activities such as commissioning, training, and planning, which are the responsibility of other Government and NHS bodies, such as NHS Digital.
- 6. CPRD operates a response mode service, whereby research organisations apply to CPRD for approval to access CPRD data to conduct their own research or commission research to be conducted by CPRD staff on their behalf. Apart from studies to characterise the quality, completeness, and representativeness of CPRD data, CPRD staff do not carry out independent research using CPRD data. All research applications to use CPRD data are subject to a rigorous governance and approvals process.

CPRD data collection

- 7. The Inquiry asks how CPRD holds data, collects data, its sources, its definitions of categories of data and how it approaches data that falls outside of these categories. CPRD research data services are based on anonymised primary care electronic health records (EHR) collected from GP practices across the UK who elect to share their data with CPRD for the purposes of public health research. No identifiable patient information (i.e.no name, address, NHS number, full date of birth or free text medical notes), is collected by CPRD. Only structured coded data from the EHR is collected.
- 8. The following categories of coded data are collected: demographic; symptoms; diagnoses; referrals; immunisations; laboratory tests; and prescriptions. While research can be carried out on individual patient-level data, these data are

anonymised and therefore it is not possible to know the identity of any patients within the CPRD database.

9. In recent years, CPRD data collected from GP practices in England have been linked to a range of secondary datasets via NHS Digital, the statutory body in England permitted to carry out data linkage processing. The data custodians of these datasets are NHS Digital (Hospital Episode Statistics), Office of National Statistics (death registry data) and Public Health England (cancer registry data).

CPRD databases and Read codes

- 10. Although CPRD data are referred to as a single database, in practice there are two CPRD databases:
 - a. CPRD GOLD, containing EHR collected from GP practices using Vision GP software; and
 - b. CPRD Aurum, containing EHR collected from GP practices using EMIS GP software.

These data are available to researchers as two separate databases, that are updated every month.

- 11. Read codes are mainly used to record medical information in these databases. These are a coded thesaurus of clinical terms that provide a standard vocabulary for clinicians to record patient findings and procedures, in the health and social care IT systems across primary and secondary care. Read codes have been used in the NHS since 1985 and Read version 2 (v2) are available in CPRD data.
- 12. The Read code hierarchy consists of over 100,000 preferred terms and about 150,000 synonymous codes, however only a subset of these codes has been used by GPs in the data collected by CPRD (between 98,000 and 106,000 codes as of June 2021). Data dictionaries and metadata are made available to researchers to enable them to interpret CPRD data, and for researchers to make assumptions and describe possible limitations of their individual studies, accordingly. Standard epidemiological approaches are used by researchers in

circumstances where information required for a study is not captured within the CPRD data.

Limitations of CPRD data

13. The Inquiry asks me to set out any limitations of the dataset held by CPRD and anything else which may affect the data, not specified, for example, where there is a lack of data.

Size of the CPRD database

- 14. CPRD was set up to be a research database of longitudinal EHR including patients who were representative of the diversity of the UK the population. For research purposes, the key value of CPRD data was the longitudinal and representative nature of the data. As I described at paragraph 7 above, the CPRD's data is collected from GP practices across the UK who elect to share their data with CPRD for the purposes of public health research. This means it is not a comprehensive holding of data from all GPs' surgeries. For many decades, the CPRD database was purposefully maintained at a relativity small size, covering less than 10% of the UK population. This figure was considered by the epidemiological research community at the time, to be appropriate for the vast majority of studies that were carried out.
- 15. In recent years, it has become a priority to grow the size of the CPRD database to develop capacity to support data-enabled clinical trials. The percentage coverage of the UK population in the CPRD database has consequently risen, e.g. from 8% in 2015 to 25% in 2021.

Changing population coverage of data over time

16. When data are first collected from a GP practice that decides to share data with CPRD, the collection includes information on patients who are currently registered, as well as a one-off collection of data from historic patients who may be deceased or no longer registered at the practice. Therefore, as the CPRD database has grown since 2015, the retrospective coverage of the population

has increased, even though the actual data volumes collected by CPRD in the years prior to 2015, may have been comparatively low.

Adoption of EHR by GP practices

17. The CPRD database was established in the late 1980s at a time when GP practices first started to convert their paper based medical records to EHR. As EHR were adopted by GPs in the subsequent decades, some practices digitised patients' partial or full records and others only collected electronic data prospectively. As a result, the completeness of EHR varies across GP practices even after 2000s, when use of EHR in primary care became almost universal. This means that CPRD data collected from GP practices covering the period of the Inquiry may not be complete.

Information captured in the EHR

- 18. Blood transfusions are likely to take place in the secondary care setting with the clinical record communicated to the GP practice via a discharge letter. GPs may add this information to the patient's EHR as a scanned PDF. In some but not every instance, information from the discharge letter may also be entered as coded data in the EHR, which is collected by CPRD. Data that are attached to patients' records in the form of scanned letters only, are not collected by the CPRD. This means that information within the CPRD data on blood transfusions may not be complete.
- 19. Laboratory test requests at GP practices only became automated from about 2003 onward, so that the recording of data relating to testing for HBV, HCV, and/or HIV infection (both test requests and results) are likely to be highly variable across the period of interest to the Inquiry. Also, laboratory test results from secondary care at this time e.g. during hospital admissions or from other community services such as sexual health or drug services, may not have been reliably communicated to the GP practice. Therefore, records relating to HBV, HCV, and HIV infection over the period of interest to the Inquiry will not be complete.

- 20. Testing for some infections such as HIV are likely to be under-reported at GP practices during the period in which the Inquiry is interested. In a published study using CPRD data (General Practice Research Database (GPRD) as was at the time), Evans and colleagues concluded that recording of HIV positive status by GPs between 1995–2005 was low and GPs may have been unaware of HIV-related morbidity amongst their patient population. (Evans HE *et al.* Trends in HIV testing and recording of HIV status in the UK primary care setting: a retrospective cohort study 1995–2005: *Sex Transm Infect.* (2009) Dec;85(7):520-6 doi: 10.1136/sti.2008.034801).
- 21. As data in the CPRD databases are anonymised, it is not possible to know the relationships between individuals, or any direct or indirect contacts individuals may have with each other. It is not therefore possible to use CPRD data to measure disease transmission.

Collection of Read Codes

22. As stated above, CPRD data includes only a subset of all possible Read codes in the Read code hierarchy. This means that searches conducted by CPRD to inform this Inquiry may not reflect or include all possible codes available in the Read code hierarchy which would provide information on HBV, HCV and/or HIV infection through blood transfusion or the use of blood products.

Data used to respond to the Inquiry

23. The data used to respond to the Inquiry are based on the June 2021 database builds of CPRD GOLD and CPRD Aurum, which represents 25% of the current UK population and data collected from a total of 60 million patients since CPRD was established. A breakdown of the numbers of patients in the June 2021 CPRD GOLD and CPRD Aurum databases, with the retrospective percentage population coverage over different time points is provided in the Table 1 below. The numbers of patients in these tables are much larger than would have been available from CPRD, or CPRD's predecessor the General Practice Research Database (GPRD), over the time the Inquiry covers.

- 24. The population coverage of CPRD for 1985, 1995 and 2005 was estimated as the number of patients registered at mid-year (01/07/YYYY) in CPRD GOLD and CPRD Aurum respectively, divided by the mid-year population estimate for the UK or England for CPRD GOLD and CPRD Aurum respectively. As shown below, population coverage was higher for the later years of interest to the Inquiry and is a direct result of increasing the number of GP practices contributing data to CPRD.
- 25. The following tables demonstrate the increasing population coverage across the two databases:

	CPRD	Mid-year UK	Percent
CPRD GOLD June 2021 build	GOLD	Population	(%)
Number of patients in CPRD GOLD:			
20,340,046			
Number of GP practices: 968			
Patients registered at mid-year (on or before			
01/07/1985) and have a registration end			
date on or after 01/07/1985	2,655,483	56,554,000	4.70
Patients registered at mid-year (on or before			
01/07/1995) and have a registration end			
date on or after 01/07/1995	5,591,833	58,024,800	9.64
Patients registered at mid-year (on or before			
01/07/2005) and have a registration end			
date on or after 01/07/2005	6,888,889	60,413,300	11.40

Table 1 – Population coverage of CPRD GOLD for 1985, 1995 and 2005.

Table 2 – Population coverage of CPRD Aurum for 1985, 1995 and 2005

	CPRD	Mid-year England	Percent
CPRD Aurum June 2021 build	Aurum	Population	(%)
Number of patients in CPRD Aurum:			
34,080,598			
Number of GP practices: 1,285			
Patients registered at mid-year (on or before			
01/07/1985) and have a registration end			
date on or after 01/07/1985	2,795,260	47,057,359.0	5.94
Patients registered at mid-year (on or before			
01/07/1995) and have a registration end			
date on or after 01/07/1995	6,938,339	48,383,461	14.34
Patients registered at mid-year (on or before			
01/07/2005) and have a registration end			
date on or after 01/07/2005	9,373,923	50606034	18.51

Section 2: Data

26. The following information has been gathered with help from my colleagues who have conducted searches across our databases. I have given the answers to the questions asked in the Rule 9 within the body of the statement and then set out the methodology for searches in the Annex, Appendices and exhibits to this statement.

Number of people directly infected with HBV, HCV and/or HIV through the use of blood transfusions and blood products

- 27. The Inquiry requests a record of the number of people directly infected with HBV, HCV and/or HIV respectively, through the use of blood transfusions and blood products. The Inquiry has explained that 'Blood products' should be taken to include products manufactured, whether by commercial or state-owned entities, from blood cells and blood plasma, including immunoglobulins. The analysis we have carried out did not find any records of individuals in CPRD GOLD or CPRD Aurum who had been directly infected with HBV, HCV, and/or HIV respectively, through the use of blood transfusions and blood products.
- 28. This is, very obviously, not to say that no individuals were directly infected, simply that the way CPRD holds data (anonymised data only that are coded and recorded by practice staff in the EHR) and searches these data (using the clinical codes recorded in the EHR) meant that no records could be identified. The reason for this is that CPRD was unable to identify diagnostic codes explicitly stating HBV, HCV, and/or HIV infection due to blood transfusions or the use of blood products, or similar statements, recorded in the database. It is important to note that CPRD diagnostic code lists are based on the diagnostic and product codes that appear in the primary care data that it collects. These codes may only represent a subset of all diagnostic codes in the respective clinical coding frame.
- 29. The methodology and search strategy criteria used are set out at the **Annex** below.

Number of recipients of blood transfusions

- 30. The Inquiry requests a record of the number of recipients of blood transfusions, either cumulatively or over a given period. I have considered the number of recipients of blood transfusions for the period 01/01/1970 31/12/2005 inclusive.
- 31.Across CPRD GOLD, the baseline population for the analysis is as follows: Number of patients born on or before 2005 and registered at a CPRD GOLD contributing practice for at least one day on or after 01/01/1970: **18,788,795.**
- 32. Please note that this analysis includes **5,568,272** individuals who were born on or before 2005 with a first practice registration date in the CPRD GOLD between 01/01/2006 31/05/2021. These individuals were included in the analysis as they may have a historical record of blood transfusion or use of blood products that may be relevant to the Inquiry.
- 33.Across CPRD GOLD, we hold the following data on the number of blood transfusions:
 - a. Number of recipients of blood transfusions recorded between 01/01/1970 31/12/2005 inclusive: **38,025**;
 - b. Number of records of blood transfusions between 01/01/1970 31/12/2005 inclusive: 61,624.
- 34. The search strategy for identifying codes for transfusions in CPRD are included in Appendix 1A. Code lists for identifying blood transfusions in CPRD are provided in 'Medical code list for identifying transfusions in CPRD GOLD' [Exhibit WITN6886002].
- 35.Across CPRD Aurum, the baseline population for the analysis is as follows: Number of patients born on or before 2005 and registered at a CPRD Aurum contributing practice for at least one day on or after 01/01/1970: **30,732,337.**

- 36. Again, please note that this analysis includes **13,420,733** individuals who were born on or before 2005 with a first practice registration date in the CPRD Aurum between 01/01/2006 31/05/2021. These individuals were included in the analysis as they may have a historical record of blood transfusion or use of blood products that may be relevant to the Inquiry
- 37.Across CPRD Aurum, we hold the following data on the number of blood transfusions:
 - a. Number of recipients of blood transfusions recorded between 01/01/1970 31/12/2005 inclusive: 45,591;
 - b. Number of records of blood transfusions between 01/01/1970 31/12/2005 inclusive: **65,779.**
- 38. Combining findings from both CPRD GOLD and CPRD Aurum yields a total of:
 - a. **83,616** recipients of blood transfusions between 01/01/1970 31/12/2005.
 - b. 127,403 blood transfusions recorded between 01/01/1970 31/12/2005.
- 39. The search strategy for identifying codes for transfusions in CPRD Aurum are included in **Appendix 1B**. Code lists for identifying transfusions in CPRD Aurum are provided in 'Medical code list for identifying transfusions in CPRD Aurum' [Exhibit WITN6886003].

The number of recipients of blood products

- 40. The Inquiry requests a record of the number of recipients of blood products, either cumulatively or over a given period. I have considered the number of recipients of blood products for the period 01/01/1970 31/12/2005 inclusive and have also provided the data for patients with a record for blood transfusions or a prescription for blood products.
- 41. The baseline populations are the same as those set out at paragraphs 31 and35 above. Again, please note that for CPRD GOLD, the analysis includes

5,568,272 individuals who were born on or before 2005 with a first practice registration date in the CPRD GOLD between 01/01/2006 - 31/05/2021. For CPRD Aurum, the analysis includes **13,420,733** individuals who were born on or before 2005 with a first practice registration date in the CPRD Aurum between 01/01/2006 - 31/05/2021. As explained above, these individuals were included in the analysis as they may have a historical record of blood transfusion or use of blood products that may be relevant to the Inquiry.

- 42. Across CPRD GOLD, we hold the following data on the number of recipients of blood products for the period of 01/01/1970 31/12/2005 inclusive:
 - a. Number of recipients of blood products recorded between 01/01/1970 -31/12/2005 inclusive: 164,523;
 - Number of records for blood products between 01/01/1970 31/12/2005 inclusive events: 212,634.

This count includes patients with a prescription for blood products (again, these are products containing blood cells or blood plasma and including immunoglobulin).

- 43. The search strategy for identifying codes for blood products in CPRD GOLD are included in **Appendix 1A**. Code lists for identifying blood products in CPRD GOLD are provided in 'Product code list for identifying blood products in CPRD GOLD' [Exhibit WITN6886004].
- 44.Across CPRD GOLD, we hold the following data on the patients with a record for blood transfusions OR prescription for blood products for the period of 01/01/1970 - 31/12/2005 inclusive:
 - Number of recipients of a blood transfusion or blood products between 01/01/1970 - 31/12/2005 inclusive: 201,850
 - Number of records for blood transfusions or blood products recorded between 01/01/1970 - 31/12/2005 inclusive: 274,258

This count includes patients with a record for blood transfusions OR prescription for blood products (again, products containing blood cells or blood plasma, including immunoglobins).

- 45. The search strategy for identifying codes for transfusions or blood products in CPRD GOLD are included in Appendix 1A. Code lists for identifying blood transfusions and blood products in CPRD GOLD are provided in the document titled, 'Medical code list for identifying transfusions in CPRD GOLD' [Exhibit WITN6886002] and 'Product code list for identifying blood products in CPRD GOLD' [Exhibit WITN6886004] respectively.
- 46. Across CPRD Aurum, we hold the following data on the number of recipients of blood products for the period of 01/01/1970- 31/12/2005 inclusive:
 - a. Number of recipients of blood products recorded between 01/01/1970 -31/12/2005 inclusive: 333,146;
 - Number of records for blood products between 01/01/1970 31/12/2005 inclusive events: 453,102.

These counts also include patients with a prescription for blood products (again, these are products containing blood cells or blood plasma and including immunoglobulin).

- 47. The search strategy for identifying codes for blood products in CPRD are included in **Appendix 1B**. Code lists for identifying blood products in CPRD Aurum are provided in 'Product code list for identifying blood products in CPRD Aurum' [Exhibit WITN6886005] and 'Medical code list for identifying blood products in CPRD Aurum' [Exhibit WITN6886006].
- 48. Across CPRD Aurum, we hold the following data on the patients with a record for blood transfusions OR prescription for blood products for the period of 01/01/1970- 31/12/2005 inclusive:
 - a. Number of recipients of a blood transfusion or blood products between 01/01/1970- 31/12/2005 inclusive: 377,884
 - b. Number of records for blood transfusions or blood products recorded between 01/01/1970- 31/12/2005 inclusive: 518,881

These counts include patients with a record for blood transfusions OR prescription for blood products (again, products containing blood cells or blood plasma, including immunoglobins).

49. The search strategy for identifying codes for blood transfusions or blood products in CPRD are included in Appendix 1B. Code lists for identifying blood transfusions and blood products in CPRD Aurum are provided in 'Medical code list for identifying transfusions in CPRD Aurum' [Exhibit WITN6886003], 'Product code list for identifying blood products in CPRD Aurum' [Exhibit WITN6886005] and 'Medical code list for identifying blood products in CPRD Aurum' [Exhibit WITN6886005] Aurum' [Exhibit WITN6886006] respectively.

Survivorship of any individuals infected with HBV, HCV, and/or HIV through blood transfusions or blood products

50. The Inquiry requests a record of the survivorship of any individuals infected with HBV, HCV and/or HIV through blood transfusions or blood products. The survivorship of individuals infected with HBV, HCV, and/or HIV through blood transfusions or blood products could not be evaluated, because as stated above, no individuals in CPRD GOLD or CPRD Aurum were identified as directly infected with HBV, HCV, and/or HIV through blood transfusions or the use of blood products.

Clinical records with a Read code that relates to blood transfusions and/or blood products by year of date associated with that event ('event date') in respect of the number of recipients of blood transfusions and blood products

- 51. In respect of both:
 - (a) The number of recipients of blood transfusions, either cumulatively or over a given period (see paragraphs 30 - 39, above); and
 - (b) The number of recipients of blood products, either cumulatively or over a given period (see paragraphs 40 - 49, above),

the Inquiry asks how many clinical records have a Read code that relates to blood transfusions and/or blood products by year of date associated with that event ('event date').

- 52. Blood transfusions and blood product records identified in CPRD may represent a subset of all transfusions and use of blood products among the search population, for reasons outlined above in Limitations of CPRD data.
- 53. CPRD GOLD: A total of 274,258 blood transfusions or blood products were recorded in CPRD GOLD during 01/01/1970 31/12/2005. About 77% of all records were for blood products which comprise primarily immunoglobulins (see 'Counts of transfusions and blood products in CPRD GOLD by year of the event from 01/01/1970 31/12/2005' [Exhibit WITN6886007]). About 95% of all documented blood transfusions or blood products were recorded during 1990 -2005 inclusive.
- 54. Of the 61,624 blood transfusions documented in the database, 300 records were documented in 1970 rising to 723 in 1989 and peaking at 5,124 records in 2004. About 95% of all documented blood transfusions were recorded between 1980 2005 inclusive.
- 55. Of the 212,634 records for blood products documented between 1985 2005 inclusive, <5 were recorded in 1985 rising to 6,193 records in 1989; most blood products were documented in 1993 (36,869 records). About 95% of all documented blood products were recorded during 1989 -1998 inclusive.
- 56. CPRD Aurum: A total of 518,881 blood transfusions or blood products were recorded in CPRD Aurum during 01/01/1970 31/12/2005. About 87% of all records were for blood products which comprise primarily immunoglobulins ('Counts of transfusions and blood products in CPRD Aurum by year of the event from 01/01/1970 31/12/2005' [Exhibit WITN6886008]). About 95% of all documented blood transfusions or blood products were recorded during 1990-2005 inclusive.
- 57. Of the 65,779 blood transfusions recorded in the database, 416 were recorded in 1970 rising to 827 in 1989 and peaking at 9,601 in 2004. About 95% of all documented blood transfusions were recorded between 1978 2005 inclusive. Of the 453,102 records for blood products documented between 1970 2005 inclusive, <10 was recorded in 1970, increasing to about 7,500 records in 1989

and peaking in 1993 (63,719 records). About 95% of all documented blood products were recorded during 1990-2005 inclusive.

- 58. 'Counts of transfusions and blood products in CPRD GOLD by year of the event from 01/01/1970 - 31/12/2005' [Exhibit WITN6886007] contains the counts of transfusions and blood products in CPRD GOLD by year of the event from 01/01/1970 - 31/12/2005. Please note that blood products were identified using prescription codes in CPRD GOLD.
- 59. 'Counts of transfusions and blood products in CPRD Aurum by year of the event from 01/01/1970 - 31/12/2005' [Exhibit WITN6886008] contains the counts of transfusions and blood products in CPRD Aurum by year of the event from 01/01/1970 - 31/12/2005. Please note that blood products in CPRD Aurum were identified using a combination of Read codes and prescription codes.

Clinical records which have at least one entry responsive to infection

- 60. The Inquiry asks about clinical records which have at least one entry responsive to infection with HBV, HCV and/or HIV respectively through the use of blood transfusions and blood products <u>and also</u> a Read code that relates to:
 - a. Testing for Hepatitis B virus, Hepatis C virus and/or HIV, by: (i) year associated with the event date for the infection, and (ii) region;
 - b. Infection with Hepatitis B virus, Hepatitis C virus and/or HIV, by (i) year associated with the event date for the infection, and (ii) region;
 - c. Any clinical records, which have an entry responsive above, which also contain the patients' date of death;
 - d. The region, country or other geographical area within the UK where the infection either occurred or was detected.
- 61. These could not be evaluated as no individuals were identified in CPRD GOLD or CPRD Aurum as directly infected with HBV, HCV, and/or HIV through blood transfusions or the use of blood products.
- 62. Please note that regardless of the absence of records identifying any individuals directly infected with HBV, HCV, and/or HIV through blood transfusions or the

use of blood products, it is not possible to determine where a patient's infection has occurred using CPRD data. It is only possible to determine the geographical area of the GP practice where the record of this infection was documented.

- 63.1 am not aware of any other categories that may assist in understanding the extent and nature of direct infection of people with HBV, HCV, and/or HIV as a result of the use of blood or blood products within the United Kingdom.
- 64. Because no individuals were identified in CPRD GOLD or CPRD Aurum as directly infected with HBV, HCV, and/or HIV through blood transfusions or the use of blood products, CPRD has not needed to account for false positives or false negatives.

Estimates of the prevalence or scale of infection in relation to HBV, HCV, and/or HIV

- 65. The Inquiry also seeks information in the following categories:
 - Estimates of prevalence or scale of infection in relation to HBV, HCV or HIV,
 - (b) Estimates of transmission rates in relation to HBV, HCV and HIV respectively.

Estimates of prevalence or scale of infection in relation to HBV, HCV or HIV

- 66.CPRD's remit is to enable public health research. As such, CPRD does not generate statistics on disease prevalence using the data that it collects as part of its day-to-day services. Researchers may access CPRD data to investigate questions on disease prevalence, the output of which may be published and available in the scientific literature.
- 67. Estimates that might be of value to this Inquiry may be available from national surveillance systems, such as that maintained by Public Health England.

- 68.1 have exhibited the following published research using CPRD data of which I am aware that may be relevant to this Inquiry:
 - a. Ferreira, GLC *et al.* Incidence and prevalence of hepatitis B in patients with diabetes mellitus in the UK: A population-based cohort study using the UK Clinical Practice Research Datalink *J Viral Hepat* (2018); May;25(5):571-580. doi: 10.1111/jvh.12841;
 - b. Gompels M, *et al.* Trends in HIV testing in the UK primary care setting: a 15-year retrospective cohort study from 2000 to 2015 *BMJ Open* (2019); 9:e027744. doi:10.1136/ bmjopen-2018-027744.

Estimates of transmission rates in relation to HBV, HCV and HIV respectively

69. As CPRD is an anonymised data source, it is not feasible to use CPRD data to address questions pertaining to disease transmission. Estimates of transmission may be available from national surveillance systems such as that maintained by Public Health England.

Section 3: Other Issues

70.1 am asked if there is anything further which has not been explicitly requested in the Inquiry's Rule 9 request that is relevant to the Inquiry. I do not believe that CPRD has further information or evidence relevant to the Inquiry's investigation of the matters set out in its Terms of Reference but we will of course assist further if there are matters arising out of this statement.

Statement of Truth

I believe that the facts stated in this witness statement are true.

	GRO-C	Y
Signed		

._....

Dated ____21/09/2021_____

Annex: Methodology used to answer the Inquiry's request for a record of the number of people directly infected with HBV, HCV and/or HIV respectively, through the use of blood transfusions and blood products

71. The CPRD search involved a search of individuals' anonymised medical records for a documented record of blood transfusions or the use of blood products. The records of these individuals were further searched for diagnostic codes and medical terms specifically indicating HBV, HCV, and/or HIV infection through blood transfusions or the use of blood products, or similar statements. The selection and use of diagnostic codes with specific medical terms indicating HBV, HCV, and/or HIV infection through blood transfusions or the use of blood products or the use of blood products by clinical staff in general practice was considered to be the most reliable search for identifying individuals directly infected with HBV, HCV, and/or HIV through blood transfusions or the use of blood products.

Search Strategy Criteria

- 72. Individuals born on or before 2005 with a record of blood transfusions or the use of blood products documented between 01/01/1970 31/12/2005 inclusive, were included. Individuals were also required to have at least one diagnostic code indicating HBV, HCV, and/or HIV infection due to blood transfusions or use of blood products in their record at any time.
- 73. Diagnostics codes indicative of blood transfusions, and of HBV, HCV, and/or HIV infection due to transfusion or the use of blood products were largely based on the Read code hierarchy, the clinical terminology used in CPRD GOLD and CPRD Aurum primary care databases. Diagnostic codes created locally by clinicians at practices contributing data to CPRD Aurum (EMIS practices) were also included.
- 74. The use of blood products, as defined in the Rule 9 Request as including 'products manufactured, whether by commercial or state-owned entities, from blood cells and blood plasma, including immunoglobulins', was determined using product codes in CPRD GOLD and medical and product codes in CPRD

Aurum. CPRD product codes are based on the NHS Dictionary of Medicines and Devices.

75. The search strategy for identifying diagnostic codes for blood transfusions, blood products and HBV, HCV, and/or HIV infection due to blood transfusion or the use of blood products in CPRD GOLD and CPRD Aurum are included in Appendix 1A &1B respectively. Medical code lists for identifying blood transfusions events in CPRD GOLD and CPRD Aurum are provided in 'Medical code list for identifying transfusions in CPRD GOLD' [Exhibit WITN6886002] and 'Medical code list for identifying transfusions in CPRD Aurum' [Exhibit WITN6886003] respectively. Product codes for identifying prescribed blood products in CPRD GOLD are provided in 'Product code list for identifying blood products in CPRD GOLD' [Exhibit WITN6886004]. Product and medical code lists for identifying prescribed blood products in CPRD GOLD' [Exhibit WITN6886004]. Product and medical code lists for identifying prescribed blood products in CPRD Aurum are provided in 'Product code list for identifying blood products in CPRD GOLD' [Exhibit WITN6886004]. Product and medical code lists for identifying prescribed blood products in CPRD Aurum are provided in 'Product code list for identifying prescribed blood products in CPRD Aurum' [Exhibit WITN6886005] and 'Medical code list for identifying blood products in CPRD Aurum' [Exhibit WITN6886005] and 'Medical code list for identifying blood products in CPRD Aurum' [Exhibit WITN6886005] and 'Medical code list for identifying blood products in CPRD Aurum' [Exhibit WITN6886005] and 'Medical code list for identifying blood products in CPRD Aurum' [Exhibit WITN6886006] respectively.

APPENDICES FOR SEARCHES CONDUCTED IN CPRD GOLD AND CPRD AURUM

Appendix 1A: Search terms used to identify transfusions, blood products and HBV, HCV or HIV infection in the CPRD GOLD database

Transfusion	Blood Products	Infection and Transfusion
Read term search	Product name	Read term search
transfu	*blood*	*hcv*transfu*
	immuno	*hbv*transfu*
Read code	*plasma*	*hiv*transfu*
search	*platelet*	*immuno*transfu*
7L14*	*packed*	*hepatit*transfu*
88*		*infect*transfu*
TA3*	Drug substance name	*virus*transfu
	blood	*contam*transf*
	immuno	
	plasma	
	platelet	
	packed	

Appendix 1B: Search terms used to identify transfusions, blood products and HBV, HCV or HIV infection in the CPRD Aurum database

Transfusion	Blood Products	Infection and Transfusion
	Product name	Read term search
	blood	*hcv*transfu*
	immuno	*hbv*transfu*
	plasma	*hiv*transfu*
	platelet	*immuno*transfu*
	packed	*hepatit*transfu*
		*infect*transfu*
	Drug substance name	*virus*transfu
	blood	*contam*transf*
	immuno	
	plasma	
	platelet	
	packed	
	Term	
	blood	
	immuno	
	plasma	
	platelet	
	packed	
	Read term search	

*immunoglobulin*vials*	
*immunoglobulin*inject*	
*immunoglobulin*infusion*	
*immunoglobulin*human*	