

Witness Name: **Sarah C. Darby** Statement No: **WITN7521001** Exhibits: **n/a** Dated: 20 December 2022

INFECTED BLOOD INQUIRY

WRITTEN STATEMENT OF PROFESSOR SARAH CAROLINE DARBY

I provide this statement in response to a request under Rule 9 of the Inquiry Rules 2006 dated 16 November 2022.

I, Professor Sarah Darby, will say as follows:

Section 1: Introduction

1. Please summarise your professional qualifications and employment history, including the roles and responsibilities you held from 1986-1996. For this period, please set out your membership of any committees, associations, parties, societies or groups relevant to the Inquiry's Terms of Reference, including the nature of your involvement.

My professional qualifications and employment history are as follows:

Education

- **1970-73** BSc Hons Mathematics, Imperial College London.
- **1973-74** MSc Mathematical Statistics, University of Birmingham.
- **1974-77** PhD Medical Statistics, London School of Hygiene and Tropical Medicine.

Employment

- **1976-78** Lecturer in Medical Statistics, Department of Community Medicine, St Thomas' Hospital Medical School, London.
- **1978-84** Senior Scientific Officer, then Principal Scientific Officer and Group Leader Epidemiology and Statistical Studies Group, National Radiological Protection Board (NRPB) (now part of Public Health England).
- **1982-83** Visiting Scientist, Radiation Effects Research Foundation, Hiroshima, Japan, sponsored by the US National Academy of Sciences (sabbatical from NRPB).
- **1984-99** Research Scientist, then Senior Scientist, then Principal Scientist, Imperial Cancer Research Fund Cancer Epidemiology Unit, University of Oxford.
- **1999-2007** Cancer Research UK Principal Scientist and Senior Group Leader, Department of Clinical Medicine, University of Oxford.
- **2007-** Professor of Medical Statistics, Nuffield Department of Population Health, University of Oxford.

Other appointments

- **1987-** Green Templeton College: Research Fellow, then Governing Body Fellow.
- **1990-** University of Oxford, Department of Statistics: Associate Member.
- **2012-17** University of Southern Denmark: Honorary Professor in Radiotherapy.

<u>Awards</u>

- **1988** Royal Statistical Society Guy Medal in Bronze.
- **1997** Professor of Medical Statistics. Title awarded by University of Oxford Distinctions Committee.
- 2019 Elected Fellow of the Royal Society (FRS).
- 2019 Elected Fellow of the Academy of Medical Sciences (FMedSci).

Scientific Societies

- **1976-** Royal Statistical Society: Council Member (1987-93), Chairman of General Applications Section (1989 91), Chairman of Oxford Local Group (2003-5).
- **1984-** International Biometric Society: Member, Member of Council (1996-1999).

1984- Society for Social Medicine: Member.

<u>Comment</u>

My work with the UK Haemophilia Centre Directors' Organisation (UKHCDO) came about through a personal invitation from Sir Richard Doll who was Chairman of the Medical Research Council's Committee on the Epidemiology of AIDS. He invited me to analyse their data because they were, at the time, lacking in appropriate statistical expertise themselves.

Section 2: Papers regarding the timing of HIV infection in UK haemophilia patients

Please refer to the following papers: OXUH0002249_015, DHSC0006372_141, OXUH0000165_002, RLIT0000370, and HCDO0000264_095. The papers were written by you and others in association with the United Kingdom Haemophilia Doctors Organisation ("UKHCDO") and relate to the timing of HIV infection in UK haemophilia patients. As to these:

- 2. In 1995, you wrote to Nature, stating: "For many patients, stored serum samples enabled the seroconversion date to be estimated reasonably precisely. The median estimated date of seroconversion was October 1982 for severe patients... and December 1982 for moderate/mild patients" [HCDO0000264_095, ep.2]. Three citations were provided: (i) Results of a 1987 survey [DHSC0006372_141]; (ii) An article in the BMJ [OXUH0000165_002]; and (iii) An article in Statistics in Medicine [RLIT0000370]. To the best of your understanding, please explain:
 - a) The basis for describing the estimates as "reasonably precisely" made.
 - b) Any qualifications (based on current statistical practice, new information, statistical methods or other factors) that you would now make to the confidence statement. Please give reasons for your answer.

c) Whether the lack of negative test data for the majority of patients affected the confidence statement (that estimates were "reasonably precisely made").

- a. The basis for this statement has two components. The first component comprises the results of a survey conducted in 1987 by the UK Haemophilia Centre Directors' Organisation (UKHCDO), which were made available to me. These results are summarised in Tables 1-3 of document DHSC0006372_141. The second component comprises the results of two statistical analyses in which estimates were derived for each patient. These analyses, presented in OXUH0000165_002 and RLIT0000370 take different approaches and make different assumptions. Despite these differing methodologies, the resulting estimates were very similar, as described in the section headed "COMPARISON OF RESULTS WITH PREVIOUS ESTIMATES FOR UK HAEMOPHILIACS" on page 687 of RLIT0000370.
- b. I am not aware of any qualifications that I would now make to this statement.
- c. The lack of negative test data for the majority of patients had little effect on the statement. This is confirmed by a sensitivity analysis described on page 687 of **RLIT0000370** where, towards the end of the last paragraph of the section "ASSOCIATION OF LATENT PERIOD WITH TYPE AND SEVERITY OF HAEMOPHILIA AND AGE AT SEROCONVERSION" it is stated "...further analyses were carried out ...in which the dates of any seronegative tests were ignored, and the last known seronegative date was taken to be January 1979 for all individuals. The results of these analyses were very similar to those taking the known seronegative dates into account."
- 3. The BMJ article stated that: "if no seronegative test was available, it was assumed that the patient was seronegative on 1 January 1979." Please explain:
 - a) The basis for this assumption. In your answer, please refer to Table 3 in OXUH0002249_015, ep.18, which indicates that

a seronegative result was only available for a minority of patients (341 of 1,201) and which indicates that large numbers of patients were seronegative in later years (e.g. 1983-1985), whereas only 4 patients were seronegative in early 1979 (the time at which those for whom no negative result was available were assumed to have been last negative).

- b) What effect, if any, this assumption will have had on the final estimated median dates of seroconversion reported in Nature in 1995 [HCDO0000264_095, ep.2].
- c) Table 3 of DHSC0006372 appears to indicate that, for patients for whom a negative test result was available, the median date of last negative test (170th patient) was in the period July-December 1982. Please explain why all patients for whom no negative result was available were assumed to be last negative at 1 December 1979 - and not later - given that the median date of last negative test was substantially later in patients for whom data was available.
- d) Please describe, in broad terms, the effect on the final estimated median date of infection if it had been assumed that patients with no negative result were last negative on a range of dates, distributed similarly to the results in Table 3 of DHSC0006372.
- a. The document OXUH0002249_015 is not a published document. Rather, it is an unpublished manuscript containing provisional and preliminary results, which was made available on a confidential basis to the members of the UKHCDO. It is dated 20/10/1988 in your request which is considerably earlier than the dates of the published documents, which are all in 1989 or later. In my opinion, it should not be considered in this discussion. Rather the finalised, published versions of the results should be considered. In this case, the relevant table is Table 3 of document DHSC0006372_141 which, as it happens, is identical to the preliminary version in document OXUH0002249_015. As explained in

5

document **DHSC0006372_141** and summarised in Table 2 of that document, the date of the earliest seropositive test in a person with haemophilia was December 1979. Thus, there was nothing to suggest that any person with haemophilia had been infected before 1 January 1979.

- b. The final estimated median dates will not be very sensitive to this assumption, i.e. if an earlier or later date (up to December 1979) had been chosen, then the median estimated dates of seroconversion as published in HCDO0000264_095 in 1995 would have been very similar.
- c. The supplementary question reads: "Please explain why all patients for whom no negative result was available were assumed to be last negative at 1 December 1979 - and not later - given that the median date of last negative test was substantially later in patients for whom data was available." However such an assumption was not made. The assumption made was that all patients for whom no negative test was available were negative on 1 January 1979, not that they were last negative on that date or on any other particular date. Infection may have occurred many months or years after 1 January 1979.
- d. The Supplementary questions asks: "Please describe, in broad terms, the effect on the final estimated median date of infection if it had been assumed that patients with no negative result were last negative on a range of dates, distributed similarly to the results in Table 3 of DHSC0006372." Please refer to page 687 of RLIT0000370 where, towards the beginning of the last paragraph of the section "ASSOCIATION OF LATENT PERIOD WITH TYPE AND SEVERITY OF HAEMOPHILIA AND AGE AT SEROCONVERSION" is stated: "Further analysis of the data showed that there were also highly significant differences between the proportions of patients developing AIDS at any given time after seroconversion according to whether or not there was an explicit date available on which a test had been carried out and found to be seronegative, with those patients for whom an explicit date was available having a higher risk of AIDS than those who did not. The majority of seronegative tests referred to blood samples obtained before 1985 and many of them would have been carried out retrospectively on

6

stored frozen serum samples. After discussion with the haemophilia centres concerned, we concluded that the most likely cause of this apparently puzzling result is that the haemophilia centres were more likely to carry out testing on retrospective samples for those patients presenting with AIDS or AIDS related conditions than for other patients." Therefore, if it were hypothetically assumed that the patients with no negative test had last negative tests with a distribution similar to that of the observed last negative tests, then the median date of seroconversion would reflect that of patients developing AIDS or AIDS related symptoms before the end of the follow-up period rather than that of all infected patients. I cannot say what the effect on the final estimated median date of infection would be under such an assumption. Please note, also, that in the question the phrase "were last negative" can be misleading. No information is available as to when any patient was last negative. Rather information is available regarding the date on which the last blood sample showing a negative result was drawn. The date of infection may have been much later than this.

- 4. Please refer to OXUH0000165_002, ep.2, which states that "the probability of seroconversion per unit time was in the ratio 18:58:1" for the three periods 1 January 1979 to 1 July 1981, 1 July 1981 to 1 July 1985, and 1 July 1985 to 31 December 1987.
 - a) Please explain the basis for these ratios.
 - b) OXUH0002249_015, ep.4 provides different ratios (8:50:1) for the same patients and time-periods. To the best of your understanding, please explain the basis for the change in ratios.
 - c) Please explain whether consideration was given to varying the probability of infection during the period July 1980 to July 1985, given changing treatment patterns in that period.
 - a. The basis for the ratios is the cumulative distribution of the percentage of people with haemophilia who were infected with HIV, given that they were eventually infected. This distribution is illustrated in Figure 1 of

7

RLIT0000370. The form of this cumulative distribution suggests that the probability of being infected per unit time (given that a person was eventually infected), was constant during July 1981 and July 1985 and lower during the periods 1 Jan 1979 to 1 July 1981 and 1 July 1985-31 December 1987. The values of ratios were then estimated assuming this form for the cumulative distribution, as described in the paragraph in which Figure 1 is embedded.

- b. As explained in my response to question 2.a above, document OXUH0002249_015 is a preliminary manuscript containing provisional results. In this document, it was provisionally assumed that the probability of infection was constant during July 1980 and July 1985 and lower during the periods beforehand and afterwards, whereas in the published document OXUH0000165_002 it was assumed that the probability of infection was constant during July 1981 and July 1985 and lower during the periods beforehand and afterwards. This different assumption regarding the change-point is the likely explanation for the different estimates. After this length of time, I do not recall the reason for the change. The value in OXUH0000165_002 should, however, be regarded as the final one and the value in OXUH0002249_015 should be ignored.
- c. The supplementary question asks: "Please explain whether consideration was given to varying the probability of infection during the period July 1980 to July 1985, given changing treatment patterns in that period." Please note that the period July 1980 to July 1985 appears only in document OXUH0002249_015, which is an unpublished manuscript containing provisional and preliminary results. In the published papers it was assumed that the probability of infection was constant during July 1981 and July 1985 (not July 1980 to July 1985). In Figure 1 of RLIT0000370, the percentage positive increases linearly during this period. This is strong evidence that the probability of infection was constant during this period. Given this evidence, no further consideration was given to varying the probability of infection during this period. The value given to the probability per unit time was chosen to be that which provided the best fit to the cumulative distribution shown in Figure 1.

- 5. Please explain the statistical methods used in these papers, their likely effect on the estimated median dates reported in Nature, and whether you consider the use of these methods to be statistically appropriate. In particular, please comment on:
 - a) Weibull distribution [OXUH0000165_002, ep.2].
 - b) The Peto and Turnbull method for estimating cumulative distribution of infections [OXUH0002249_015, ep.4].
 - c) Any other methods you consider had a material impact on the final estimates.
 - a. As explained in 3.a above, the form of the cumulative distribution of infections and the locations of the cutpoints was suggested by Figure 1 of **RLIT0000370**. However, an additional assumption about the distribution of the latent period (i.e. the time between seroconversion and the development of AIDS) was necessary in order to estimate values for the ratios of the probabilities of infection per unit time. The assumption made was that the latent periods followed a Weibull distribution. This distribution is commonly used in modelling quantities such as latent periods because of the wide variety of shapes it can take. The shape is determined by two parameters, and the values of these parameters are chosen so that the distribution follows the observed data as closely as possible.
 - b. As explained in 2.a above, document OXUH0002249_015 is a preliminary unpublished manuscript. However, the method used in OXUH0000165_002 also refers to the papers by Peto and Turnbull. The paper by Peto describes a method for estimating a cumulative distribution when some of the observations are right-censored, i.e. known to have occurred before a certain date, as with the dates of the reported dates of first seropositive tests as summarised in Table 2 of DHSC0006372_141. The method of calculation described in the Peto paper is somewhat complex and the paper by Turnbull presents a simpler method.

- c. I am not aware of any other methods having had a material impact on the final estimates that I published.
- 6. Please comment on any other aspects of these papers and the data relied on which you consider significant and/or which had a material impact on the final estimates reported in Nature in 1995.

I am not aware of any other aspects of these papers or the data provided to me by the UKHCDO as having had a material impact on the final estimates reported in *Nature* in 1995.

7. Please provide the Inquiry with any other documents or papers in your possession or knowledge which you consider significant to the issues to which this request relates. Please explain why you consider any such material to be significant.

I do not have and am not aware of any other documents or papers that are significant to the issues to which this request relates.

Statement of Truth

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

Sianed	GRO-C	
	i	

Dated	20 December 2022	
-------	------------------	--