Report to the Court of Appeal Re: Applicant Benjamin David GEEN

Prepared for: Bastian Lloyd Morris

1. I am instructed by Bastion Lloyd Morris Solicitor Advocates of 209 Sovereign Court, Witan Gate East, Central Milton Keynes, MK9 2HP, to prepare a Report concerning the case of R v GEEN. Three points were listed in the letter of instruction of 15 Sept 2009.

2. My brief answers to the three questions are:

   (a) The Crown cannot reasonably or fairly simply declare a pattern to be unusual.

   (b) Opinion evidence can be very misleading.

   (c) I regard the information presented in this case as of inadequate quality. The collection and analysis of data on cases of 'sudden collapse involving breathing difficulties (SCBD)' is such that I do not think the evidence is of any value in assessing frequency of patterns, hence it is not of value in making inferences as to causes. A statistical analysis, perhaps several are required to evaluate unusual patterns.

3. My knowledge of the case is presently limited to the following documents:

   Documents read:

   1. A letter of instruction of 15 Sept 2009.

   2. Report by Professor David Denison, dated 30th June 2009.

   3. Court of Appeal Case Summary: Regina v Benjamin David GEEN, Reference 2007/5540/D3.

   Documents received; some pages skimmed:

   4. Summing up, Day 1.

   5. Summing up, Day 2.

   6. Summing up, Day 3.

   7. Summing up, Day 4.

   8. Transcript T20047186/T20047114.

Identification and selection of cases, collection and analysis of data

4. There are standards of data collection and analysis which are expected of scientific evidence, if it is to be deemed reliable. Some common sources of inadequacy are described by Professor David Hand in Appendix D of The Law Commission Consultation Paper No 190 [1].

5. The summing-up indicates that Professor Aitkenhead said that although no records are kept which he could quote, he was well aware of how many such events occur. As Professor Aitkenhead is aware of how many events occurred, I would have appreciated some numerical substantiation of the rarity or otherwise of events. The question of what should be regarded as expert evidence when that evidence includes probability and statistical judgements has been discussed within the statistical community, and with legal views considered [2].

6. In gathering information or evidence relevant to assessing a pattern, routine data sources should be investigated. Special data collection should follow the principles of good design, data collection and analysis relevant to surveys, cohort studies and case-control studies [1, 2, 7]. To assess causation, controlled experiments are most powerful, but are not always possible, and evidence from various sources has to be weighed [7, p 10-13].

Recommended approaches

Routine data

7. Office for National Statistics data on causes of death can be reviewd to see how common respiratory arrest is as a cause of death. To illustrate the need to consider recorded evidence, rather than impressions, consider Department of Health advice targets which related to preventable deaths in hospital. Initially, government targets focused on death from hospital acquired infections of MRSA (22 July 2005)[3], which had been discussed in the media. Awareness of Clostridium difficile, and GRE increased later. However, although a fair number of deaths in hospital are from deep vein thrombosis, the Department of Health website does not have advice of the same detail as for MRSA.

8. The External Review Team found five incidents in December 2002, before Geen was present, and six in December 2003. To assess the implications of associating all six of the December 2003 cases with Geen, the implicit assumption must be made explicit and evaluated. The assumption is that the rate of 'natural cases' in December 2003 was zero, whereas it was five in December 2002. If the underlying

[1] http://www.lawcom.gov.uk/docs/cp190.pdf

[2] Aitken, Schafer and Mavridis, *Expert statistics*, personal communication, 2008

[3] http://www.dh.gov.uk/en/Publichealth/Healthprotection/Healthcareacquiredinfection/Healthcareacquiredgeneralinformation/DH4093113

rate of 'natural cases' in December were five, the probability of no natural cases in December 2003 is less than 1%, and the probability of one or fewer is 4%.

Analytical studies

9. Any cohort or case-control study needs to limit the risk of bias. Nine important causes of bias are discussed in an article *"Bias in analytical Research"* by Sackett [12], and 35 biases listed. Numerical evidence for many of the biases is provided. Some of these biases can be removed in cohort studies.

10. In this instance, *diagnostic suspicion bias (6, op.cit)* is very important. The danger is that 'A knowledge of the subject's prior exposure to a putative cause ... may influence both the intensity and the outcome of the diagnostic process.' The relevant 'diagnosis' is "sudden collapse involving breathing difficulties (as opposed to cardiac arrest) which could only be explained on the basis that something unauthorised had been injected." Knowledge that Mr Geen had been on duty might influence the effort made to find an alternative explanation for collapse. For this reason, there should be a clear statement of the possible causes of sudden collapse, and careful assessment of each of these possible causes, with the evidence for and against each cause.

11. *Prevalence-incidence bias* refers to the timing of case-finding: rapidly fatal cases, or short episodes can be missed. Were there patients who died before being admitted to 'resus'? Were there incidents of breathing difficulties which resolved, and did not lead to admission to 'resus'?

12. *Unmasking (detection signal) bias* occurs when an innocent exposure becomes suspect because, without causing a disease, it precipitates the search for a disease. So, if the 25 unexplained cases were identified by looking at the notes of all patients admitted while Mr Geen was on duty, there is a very high risk of creating an unmasking bias.

13. *Missing clinical data bias.* If the detail in the case notes, and the range of tests carried out vary between potential cases and controls, there is a risk of assuming that no record of results for a test is interpreted as a negative result, i.e. a lack of explanation for collapse, rather than as showing that a cause has not been considered.

14. *Exposure suspicion bias.* "A knowledge of the patient's disease status may influence both the intensity and outcome of a search for exposure to the putative cause". The example given is of the increase in estimate of exposure to radiation in cancer patients after intensive questioning and searching of records. Publicity about a nurse can also lead to exposure suspicion bias, particularly if cases reported to the hospital were in anyway filtered to exclude cases when Mr Geen was not on duty.

15. *Recall bias.* Questions may be asked more often of cases than controls: mothers of children with malformations are more likely to be asked detailed questions

about smoking, alcohol and other drugs. We are more likely to think carefully about our previous day's meals if we wake up feeling ill.

16. *Family information bias* refers to the effect of a case on family memory, which would also apply to memory of staff once a person is suspected of an offence. The example in the article shows substantial differences in what siblings report about parents arthritis, according to whether the child did, or did not have, arthritis.

Cohort studies

17. Cohort studies select individuals from a defined population, and then record which of them are exposed to the putative risk factor, and which develop the disease or event of interest [7, p 16-17]. Membership of the cohort is defined before the event of interest occurs. A population might be defined retrospectively, e.g. all patients admitted to A&E between 1 November 2003 to 29 February 2004, and medical notes used to assess whether the event occurred; or prospectively, with all members of the cohort from a certain date followed-up and subject to well-specified diagnostic tests and criteria. These tests and criteria should be evaluated without knowledge of the exposures of the cohort members.

18. The External Report indicates that at first 18 patients were identified as possible cases of SCBD, but on review, the number was reduced to 10. Later 25 unexplained cases were alleged; charges were preferred in 18 cases. The changes in which incidents were defined as SCBD indicate that it is likely there was not one agreed definition. In particular, it appears that at least one possible case was unexplained, and then became unexplained again. The increase in the number of possible cases after publicity suggests the possibility of exposure suspicion bias and recall bias.

19. A clear statement of how all the 18 and 25 cases were identified should be given. Were *all* patients admitted to A&E in a defined time-period were considered? Was a clear, unambiguous definition of 'sudden collapse involving breathing difficulties' documented in advanced of the search for cases? Was there a full, detailed list of possible explanations, including the methods used to find and assess evidence for and against each possible explanation? Is there a report of all decisions as to why cases were included or excluded?

20. As intervention by a member of staff was one of the hypothesised causes, a list all staff who were present during the time under study is required: all staff, not only nurses: doctors, porters, cleaners and receptionists. My reason for including doctors is that they also attend patients, and have access to drugs. My reason for including other staff is that there might be factors such as differences in approaches to triage.

21. The total times worked by staff over the index period might be relevant. Any comparisons of rated of incidents must take into account the time over which incidents were observed.

22. Document all changes in A & E practice in the six (say) months before the index period. This includes aspects such as a change in preferred cleaning materials. Breathing difficulties can arise from aerosols, and from other cleaning, building and decorating materials. Clothes strongly saturated with tobacco smoke can affect me. As a different example, a neonatology unit changed the diet for very premature babies: the death rate increased substantially.

23. Care should be exercised in creating the list of staff present. If planned rotas are used, a check should be made of whether staff missed a shift because of illness, or whether staff swapped shifts.

24. As some of the alleged cases were dropped, the statement that Geen was the only nurse present on all occasions might no longer be true. It is not clear to me whether Mr Geen was present with all 18 or 25 patients. It is possible, indeed likely, that when the number of cases is reduced from 18 to 10, a few more staff would be present on all occasions. A table of staff against all possible cases should be presented. If ten cases remained after peer review, how were the other 15 cases 'missed': were satisfactory explanations for those patients' conditions found, and then these explanations rejected after the public response? Did the public response generate extra cases which were submitted to the police?

25. The analysis of the results of the cohort should include numerical summaries of the risk of being a case associated with the different risk factors, and the putative explanations.

26. **Additional cohorts:** In order to assess whether the incidence of SCBD in Horton General Hospital was unusual, the incidence in other hospitals should be studied. Statistical and epidemiological practice is that in seeking to establish causality, consistency across populations and investigators should be considered [2, p 136], [7, p 10-13]. It would be usual to use the same time of year, to control for seasonal effects, and to study a cohort with the same design criteria as the cohort study at Horton General Hospital. As I understand that SCBD is not a simple, easily defined condition, recall by health staff of incidents over passed months is unlikely to be very accurate.

27. If after careful collection and analysis of data, a pattern can be said to be unusual in the sense of being unlikely to occur by chance, we cannot conclude that it has not occurred by chance. Competing explanations would usually be considered; often some-one would have to make a decision between chance and one alternative explanation.

Case-control studies

28. Case-control studies are generally quicker and less expensive that cohort studies [1, p 93-96]. The main dis-advantage is that case-control studies are open to more biases [12].

29. The requirement for clear definitions, and consistent search for possible exposures, are the same for a case-control study as for a cohort study.

30. A case-control study of patients who were transferred to 'resus' after admission to A&E would define a case as 'a patient who was transferred to 'resus' more than, say, 30 minutes after admission to A&E'. An obvious control population would be all patients admitted to A&E, but not transferred to 'resus', perhaps with some criteria of severity such as patients who were brought in by ambulance, or who were unable to walk in unaided.

31. A complete list of all such cases and controls for the time period of interest would be needed. It might be that one control per case were used, or more cases per control, and cases might or might not be matched [7].

32. The time period could be the three months concerned (December 2003-February 2004) and the equivalent for the preceding two years and the following year. Periods before and after index period might also be relevant; this kind of decision is usually made by a multi-disciplinary group.

33. A very important aspect of a case-control study is that information on all cases and controls must be reviewed by a person who does not know whether the patient is a case or a control. In some studies, each case and control is independently reviewed by two reviewers who are blind to the case or control status.

34. Consider competing explanations and define what would provide evidence for the various explanations in advance, so that information on possible exposures is systematically collected.

35. Provide statistical summaries and estimates of the alternative factors associated with cases.

Unusual patterns

36. There is considerable variation from hour to hour, day to day, week to week and month to month in the factors which bring patients into A& E. It is therefore necessary to consider carefully how to establish what the usual pattern is.

37. Any statement that a pattern is unusual implies a standard, or usual pattern. Four spatial plots are given in Figure 1. Which of these figures is regarded as unusual depends on what we are told, or assume.

    If asked whether there were any evidence of clustering round points, and how many clusters, most people would pick out Figure 3, and a good number would pick out Figures 1 and 4. There would be a range of views on the number of clusters. In fact, only one of figures 1, 3 and 4 has genuine clusters, that is, clusters which I know, because I created the patterns, have not arisen by chance.

38. If we are told that the four figures represent tents in a large, flat field, we might think that Figure 2 is usual for a meeting of scouts, and that Figure 3 is very unusual. If we are told that these are houses in a very mountainous region, then Figure 2 becomes unusual, and Figures 1 and 4 might be usual: Figure 1 if the points follow valleys, Figure 4 if there are steep slopes in the bottom left of the

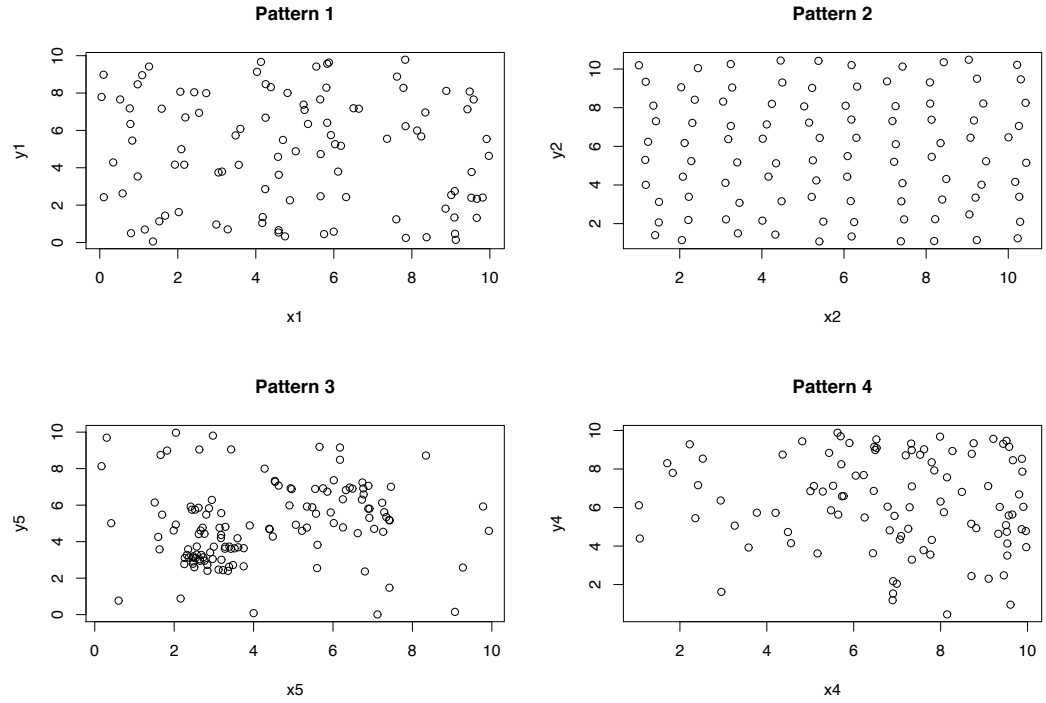

Figure 1: Spatial patterns

figure, and flatter ground towards the top right. If we are told these are trees, then Figure 1 might be usual for thorn trees on a savannah, and Figure 2 usual for a dense copse of hazel trees.

Examples of possible clusters or patterns.

39. **Corby cluster of birth defects**

A High Court ruling on 29 July 2009 included the judgement that "there was a statistically significant cluster of birth defects between 1989 and 1999." Two epidemiological experts disagreed on whether the cluster was statistically significant but the judge was most impressed by the one that concluded it was. If the rate of defects in Corby (6/7,736) is compared with the rate in the rest of Kettering Health Authority (8/27,891), the rate in Corby looks high. In considering whether this is a fair comparison, it is useful to imagine asking a group of people, say the 30 academic staff of the Warwick Statistics Department, to line up in order of height. Should one be surprised if the shortest person is a few inches smaller than the tallest? Is the shortest person too short, i.e. a dwarf?

Professor David Spiegelhalter, Winton Professor of the Public Understanding of Risk, Cambridge University, has published a commentary on the web, which discusses the issues and assumptions behind the data and the statistical analysis.

http://www.straightstatistics.org/article/question-marks-over-corby-judgement

Professor Spiegelhalter shows that clusters will arise by chance in the UK, so that there is a need to evaluate whether the Corby cluster can be explained by chance.

He points out that the ideal is to have two independent sets of data, so that the hypothesis of a cluster is tested separately. If this is not possible, we have to consider how clusters might arise. He shows that if Kettering Health Authority is divided into five districts, and 14 cases are allocated at random to these five districts, then the chance that one district has 6 cases is 14% to 22%.

40. **Sellafield Nuclear clusters of childhood leukaemia.**

    In 1984, apparent clusters of childhood lymphoid malignancy in Northern England were reported in *The Lancet*[5]. Results of a case-control study published in the *British Medical Journal* in 1990 lead to the conclusion: "The raised incidence of leukaemia, particularly, and non-Hodgkin's lymphoma among children near Sellafield was associated with paternal employment and recorded external dose of whole body penetrating radiation during work at the plant before conception. The association can explain statistically the observed geographical excess" [9, 8]. An alternative explanation, population mixing, was reported in 1999, *British Journal Cancer*; this model predicted three, rather than six, cases of acute lymphoblastic leukaemia and non-Hodgkin's lymphoma [6]. This example illustrates the importance of considering a range of possible explanations.

41. **Cluster of deaths on nurse's shift**

    Statisticians have commented on the use of statistics in the conviction of a nurse, Lucia de Berk in Holland, of murder. Lucia De Berk witnessed a high number of incidents during her shifts. An accessible account of this is provided by the journalist Ben Goldacre:

    http://www.guardian.co.uk/science/2007/apr/07/badscience.uknews

    An extended discussion of the possible statistical analysis makes the point that several models exist, and these models lead to very different predictions of the number of incidents [10].

42. **Cluster of serious events in children with cystic fibrosis.**

    (a) In 1993, doctors at Alder Hey Children's Hospital (AHCH), Liverpool, noticed that five children with cystic fibrosis (a condition in which the lungs and digestive system are clogged with thick sticky mucus) who needed surgery because of fibrosing colonopathy (obstruction of the intestine) presented between July and September, 1993. One response to this might have been to suggest that doctors at AHCH were failing in some way.

    (b) On 8 January 1994, a short report was published, which reported that "The only consistent change in management had occurred 12-15 months preciously when all five had switched to" high-strength pancreatic enzymes *(high dose drugs)* [15].

    (c) At the time the report was published, a case-control study to investigate the findings had been started: this is the appropriate method of reacting to the reports of new adverse events among patients. The Medicines Control Agency had been informed of the cases, and had issued appropriate warnings. There were about 7600 people known to have cystic fibrosis in the UK; 5/7600 is 0.07%.

(d) On 11 November 1995, about 2 years later, the results of the case-control study were published [14].

(e) The study had 14 cases of fibrosing colonopathy, with each case matched to four controls. Data on these 70 patients showed a significant (at 5%) odds ratio of 1.45 per extra 1000 high-strength capsules, and indicated which two particular proprietary formulations were associated with the highest odds ratios. That is, this association between particular formulations and fibrosing colonopathy could have arisen by chance one time in twenty.

(f) Laxative use was also found to be associated with fibrosing colonopathy; odds ratio 2.42 (95% Conf. Int 1.20-4.94). From a case-control study, one cannot establish whether laxative use was a cause of fibrosing colonopathy, or a symptom of it.

(g) Six of the 14 cases received care at AHCH. Care at Liverpool was associated with approximately a two-fold increase in risk of fibrosing colonopathy. If taken alone, this risk is statistically significant at the 4 percent level (p=0.04%), but adjusting for high-dose drugs removes the significance (p=0.3 or p=0.8).

(h) In deciding whether to suggest that AHCH doctors were negligent, or actively harming children with cystic fibrosis, one must consider the competing explanations for fibrosing colonopathy.

Risks of opinion evidence

43. Most people do not keep accurate records of events, and memory is generally selective. People complain that the other queue is faster, or buses are unreliable, or that trains are always late, particularly if they have recently been inconvenience by a late train. An important part of cognitive behavioural therapy is asking clients to keep a diary, so that they can assess whether the problems which they perceive are as severe as they think [4]

44. People often also have difficulty extracting the numerical evidence and combining it to reach the correct conclusion. Examples of the implications of screening tests of different accuracy are often used by statisticians to illustrate these difficulties [2, Ch 3]. The journalist Michael Blastland uses screening for terrorists to illustrate this point [5]: he asks "If a machine for screening for terrorists is 90% accurate, and a result is positive, how sure are you that this person is a terrorist?"

45. Medical doctors are not specifically trained to make numerical estimates, and objective recall of events is difficult for most people. Randomised controlled trials are now required for most new interventions, because bias of many kinds in estimating the benefits of treatment is well documented. Systematic reviews of randomised trials are now used by the National Institute for Clinical Excellence, in order to provide sound evidence to inform practice.

[4] http://www.rcpsych.ac.uk/mentalhealthinfoforall/treatments/cbt.aspx
[5] http://news.bbc.co.uk/1/hi/magazine/8153539.stm

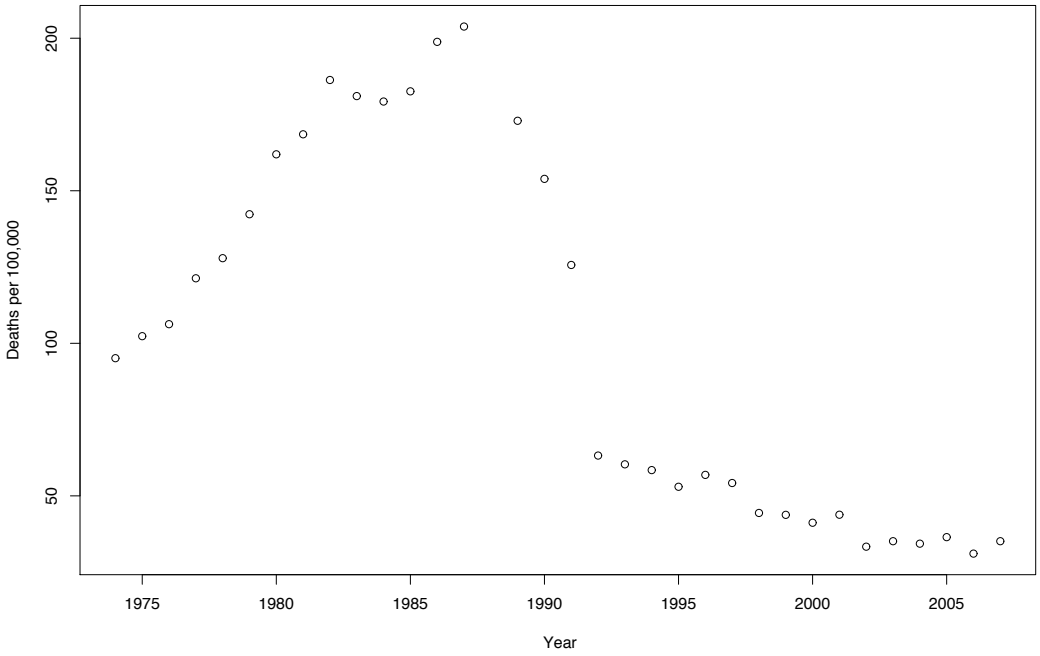

Figure 2: Death rate under 1 year: all unknown causes

46. Some tens of thousands of babies died because of medical advice to put babies to sleep on their stomachs. A very influential source of this advice was the paediatrician Dr. Benjamin Spock, through his book *The Common Sense Book of Baby and Child Care, 1946.* After case-control studies showed that medical opinion on sleeping position was wrong, advice was changed, and the death rate dropped dramatically, Figure 2.

47. Human albumin was given to critically ill patients for many years. It was a respected and widely used treatment[11]. In 1998 a systematic review was published which estimated that albumin was associated with extra mortality: for every 100 patients treated, an average of six extra deaths were caused [4, 11]. Medical opinion, given in good faith, was that human albumin was beneficial, but doctors were not able to keep an accurate mental record of the deaths of patients who did, and did not, receive albumin. The doctors did not realise they were increasing the death rate.

48. For RCTs, properly controlled trials generally have smaller treatment effect estimates. A large study showed that inadequately controlled studied exaggerate treatment effects by 41% [13].

49. Doctors have been found to be inaccurate in estimating survival for terminally ill-patients, even over time-scales of less than a month [3]: only 20% of predictions were accurate to within 33% of actual survival.

Conclusion

50. The evidence of an unusual pattern of sudden collapses given in the summing up was of no value in supporting a conclusion that there was an unusual pattern, nor a conclusion that any unusual pattern was not a chance event. Opinions given by expert or other witnesses which are based on anecdotal evidence are very likely to be misleading.

References

[1] D. G. Altman. *Practical Statistics for Medical Research*. Chapman and Hall, London, 1991.

[2] M J Campbell and D Machin. *Medical statistics: a commonsense approach*. Wiley, Chicester, 3 edition, 1999.

[3] N A Christakis and E B Lamont. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *Brit. Med. J.*, 320:469–473, 2000.

[4] Cochrane Injuries Group Albumin Reviewers. Human albumin administration in critically ill patients: systematic review of randomised controlled trials. *Brit. Med. J.*, 317:235–240, 1998.

[5] AW Craft, S Openshaw, and J Birch. Apparent clusters of childhood lymphoid malignancy in Northern England. *Lancet*, 2(8394):96–7, 1984.

[6] HO Dickinson and L Parker. Quantifying the effect of population mixing on childhood leukaemia risk: the Seascale cluster. *Br J Cancer*, pages 3–5, 1999.

[7] R Farmer, D Miller, and R Lawrenson. *Epidemiology and Public Health*. Blackwell Science, Oxford, 4 edition, 1996.

[8] MJ Gardner, AJ Hall, MP Snee, CA Powell, S Downes, and JD Terrell. Methods and basic data of case-control study of leukaemia and lymphoma among young people near sellafield nuclear plant in west cumbria. *Brit. Med. J.*, 300:429434, 1990.

[9] MJ Gardner, MP Snee, AJ Hall, CA Powell, S Downes, and JD Terrell. Results of case-control study of leukaemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria. *Brit. Med. J.*, 300:423429, 1990.

[10] R Gill, M Collins, M van Lambalgen, and R Meester. On the (ab)use of statistics in the legal case against the nurse lucia de b. *Law, Probability and Risk*, arXiv.org/math.ST/0607340, 2007.

[11] M Offringa. Excess mortality after human albumin administration in critically ill patients. *Brit. Med. J.*, 317:223–224, 1998.

[12] DL Sackett. Bias in analytical research. *J Chronic Dis*, 32:51–63, 1979.

[13] KF Schulz, I Chalmers, RJ Hayes, and DG Altman. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *J. Am. Med. Ass.*, 273:408 – 412, 1995.

[14] R L Smyth, D Ashby, U O'Hea, E Burrows, P Lewis, D van Velzen, and J A Dodge. Fibrosing colonopathy in cystic fibrosis: results of a case-control study. *Lancet*, 346:1247–1251, 1995.

[15] R L Smyth, D van Velzen, A R Smyth, D A Loud, and D P Heaf. Strictures of ascending colon in cystic fibrosis and high-strength pancreatic enzymes. *Lancet*, 343:85–86, 1994.

Declaration

1. I have set out the substance of all the instructions I have received (with written and oral questions) upon which my opinion has been sought and the materials provided and considered, and the documents, statements, evidence, information or assumptions which are material to the opinions expressed or upon which those opinions are based.

2. I have stated the facts and assumptions upon which my opinion is based.

3. I have identified, where relevant, the questions and issues lying outside my expertise.

4. Insofar as there is a range of professional opinion I have set that out together with the reasons for forming the opinion I have expressed.

5. I have set out my academic and professional qualifications relevant to the opinion I have expressed herein and the range and extent of my expertise, identifying where relevant any limitation in that expertise.

6. I believe I have complied with my duty to the Court to provide independent assistance by way of objective unbiased opinion in relation to matters which are within my expertise and I will inform the Prosecution, the Defence and the Court in the event that my opinion subsequently changes on any material issue.

Signed

Professor Jane L Hutton, September 29, 2009

Department of Statistics,
University of Warwick,
Coventry,
CV5 6AB.