

CONFIDENTIAL to CERRIE

NORDIC LEUKAEMIA STUDY

Do the Data Sets Show Any Non-Random Features?

Data Analysis by Philip Day, University of Manchester

November 2003

This paper is intended for the information of CERRIE only, on a matter which has been discussed in Committee.

SUMMARY

The Nordic Leukaemia study has been the subject of a certain amount of discussion by the Committee. The principal question that appears to have been examined is whether the incidence of leukaemia in children in the Nordic countries shows an increase that could be related to radioactive fallout from weapons testing.

In the present analysis, data sets giving leukaemia incidence annually in children of various age groups up to age 5, from 1948 to 1983, have been examined for the five Nordic countries, Denmark, Norway, Sweden, Finland and Iceland. For analysis, the data (male/female combined) have been grouped by year of birth. The annual incidence figures are small, around 5 - 30 cases to age 5 in each year group (around 6-8 cases/100,000 population). The annual incidence will be expected to fluctuate, and the question asked in this study is whether any of these fluctuations are large enough to be considered non-random, and therefore to merit further investigation as to their cause.

The conclusions reached are that, with the exception of Denmark, the data for the Nordic countries show annual variations reasonably in line with a Poisson distribution. For Denmark, there is a non-random anomaly, not replicated in the data for the other four countries. Specifically, for children under 5 years of age, born in Denmark in the years 1965 and 1966, the leukaemia incidence is far above the likely excursions of a normal statistical fluctuation. A specific physical cause seems likely.

If radioactive fallout were postulated as a likely cause of the Danish anomaly, the absence of a similar anomaly in the other four countries would require explanation. Thus, whilst a raised leukaemia incidence in the 1965/66 birth cohorts from Denmark undoubtedly has occurred, no obvious cause is apparent.

DETAILED STUDY

1. Purpose and Background

The Nordic Leukaemia study (Darby and others) has been the subject of a certain amount of discussion by the Committee. The principal issue that has been raised (Busby and others.) is whether, contrary to earlier reported findings, the occurrence of leukaemia in children in the Nordic countries showed a raised incidence over a consecutive period of years that can be related to the period of enhanced radioactive fallout from weapons testing.

The Purpose of the present study is to examine the Nordic Leukaemia data in isolation, purely to determine whether or not any of the annual variations in leukaemia incidence are outside the statistical fluctuation that would be expected in such circumstance. There is no intention to attempt to relate any of the disease data to radiation doses or any other environmental variable.

The available data cover children from the 5 Nordic countries, from birth to age 15, over the period from 1948 to 1986, although the actual time period covered for any one country does not cover the whole period. Atmospheric weapons testing largely took place in the 1950s and early 1960s, and terminated with the test ban treaty in 1964. Atmospheric fallout increased over the late 1950s and reached a peak in the Scandinavian countries in the mid-1960s. It is therefore plausible that, if an increase in leukaemia is observed, there could be a causal relationship with radioactive fallout. Previous epidemiological investigations (Darby et al.) have tested for a relationship between incidence and estimated radiation dose to various age groups, and although a small increase in leukaemia incidence over the high-fallout period was observed, no significant relationship with radiation dose could be detected. The counter-claim (Busby et al.) has been that the incidence data for Denmark particularly show an unexpected surge in leukaemias in the mid-1960s, and that it is likely that this surge was related to the increase in fallout at this time. (A brief summary of the essential data defining this controversy is given in Appendix I to this document.)

Data sets are available for the countries of Denmark, Sweden, Norway, Finland and Iceland. Most data sets cover a significant fraction of but not the whole time period, the Danish set covering the widest range of years. Leukaemia incidence is given for each 1-year age group (both male and female), as the number (a) diagnosed in any given year, and (b) by year of birth.

The principle analysis reported here is for children under 5 years of age from all five Nordic countries. The incidence of leukaemia for children under 5 has been analysed according to the year group in which they were born.

2. Analytical Method: Example

The objective of the present analysis is simply to examine each leukaemia data set in isolation, in an attempt to determine whether or not there is any apparent anomaly in the annual incidence which would merit further (epidemiological) investigation. The number of cases occurring each year is small (average <30), and the numbers would in any event be expected to fluctuate. The analysis sets out to determine whether any of the apparent fluctuations in annual incidence are large enough to be significantly outside expected statistical variability.

Formally, the hypothesis under examination is that the observed annual fluctuations in leukaemia incidence are within the range which might reasonably be ascribed to chance. The central assumption is that annual fluctuations over a period of several years will be expected to follow a Poisson probability distribution about the mean incidence over the whole period. The analysis identifies only whether there are any large deviations from this expectation, and does not set out to examine possible causes.

The method adopted is demonstrated by the following example, using data for Denmark. The number of leukaemia cases occurring before the age of 5 for each annual birth cohort is considered sequentially. The steps taken are set out below.

1. *Raw data plot.*- The data for Denmark categorised by year of birth are examined, and data relating to the age groups 0-4 years (i.e. covering the first 5 years of life, both sexes) have been extracted. The number of leukaemia cases diagnosed for each annual birth cohort are plotted in Figure 1 (upper line, **a**) for each birth year from 1948 to 1984. These average at 25.7 cases/year, although there is clearly a slight downward trend over the time period. This trend largely disappears if the incidence is related to population size; Figure 1 (lower line, **b**) plots the annual incidence normalised to a population of 100,000.

2. *Poisson distribution.*- Working with the actual number of cases/year, the Poisson distribution appropriate to the average (in this case, 25.7) can be calculated, and also the cumulative probability of the occurrence of a number n or greater. These distributions are shown in Figure 2.

3. *Probability plots.*- Figure 3 (lower line, **a**) plots the probability of occurrence of each annual observation (using the Poisson distribution shown in Figure 2). From examination of Figure 1, it appears that the number of cases recorded for children born in 1965 and 1966 (41 and 50, respectively) deviate the most in the sequence plotted, prompting the question: do these years demonstrate anything significantly outside random deviation? The probability of 50 or more cases occurring in any one year cohort, given an average value of 25.7, is calculated as 0.000014, i.e. around 1 in 70,000[†]. Thus, the occurrence of 50 cases in the 1966 birth cohort is very unlikely to have been a random fluctuation in the data set.

[†] This should not be judged in relation to the "normal" test of statistical significance ($P < 0.05$) because that test is intended to apply to single events. In this case, 37 years have been plotted and the equivalent test would be $P < 0.0014$ (i.e. $0.05/37$). The observed probability of the selected event is 100 times below this level, and one may conclude it is unlikely to have occurred as a random fluctuation.

Figure 3 (upper line, **b**) takes the analysis a stage further, and calculates the probability of the observed number of cases for each year triplet. That is, the point plotted for 1949 is the probability that the occurrences for the 1948, 1949 and 1950 cohorts would all have occurred in sequence. The basis for this analysis is the argument that if the annual incidence of leukaemia were being influenced by a specific environmental factor, then the expectation would be that cases of unusually high incidence might occur in groups because, if the outside influence lasted for a few years only, then a higher rate of incidence might be expected over the course of the few years at or after some interval following the exposure period. The plot shows clearly that the occurrence of leukaemia in the birth cohorts between 1963 and 1966 deviates markedly from the apparent norm for the remaining years. At the peak, the probability of the observed cases occurring as random fluctuations in a Poisson distribution falls to below 1 in 10,000,000.

4. *Statistical conclusions.*

- The occurrence of leukaemia in children under 5 in Denmark born in the years 1965 and 1966 is very significantly higher than the average incidence in the birth cohorts over the 1948-84 period.
- It is unlikely that this anomaly represents a chance fluctuation in the incidence rate.

FIGURE 1. Leukaemia incidence in Denmark, age group 0-4, by year of birth.

The upper line (a) shows the recorded number of cases by year of birth (the average, 26, ± 2 x standard deviation are shown as dashed lines). The lower line (b) shows the number of cases standardised to a population of 100,000.

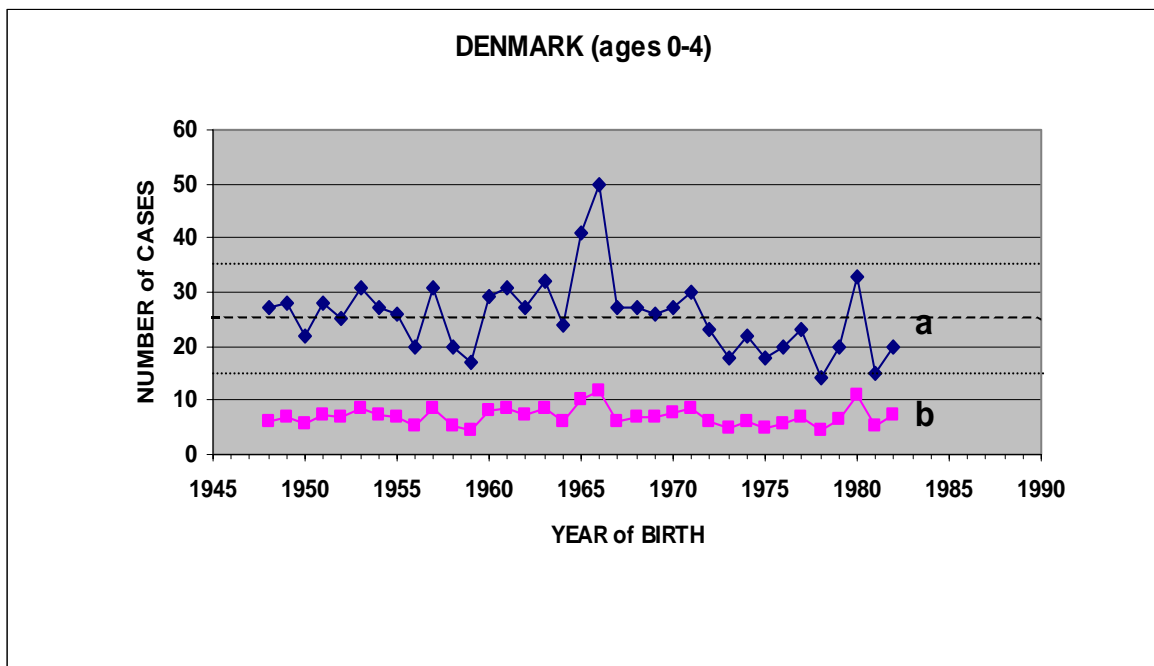


FIGURE 2. Poisson Distribution Function[‡] related to an average value of 25.7. The blue line (points as triangles) gives the probability of the occurrence of any specific value of n (for clarity, the probabilities have been multiplied by 10). The continuous curve (red) gives the cumulative probability; that is, the probability that any number equal to or greater than n will occur. For example, the probability of 26 cases occurring in any one year is 0.078, whilst the probability of 26 or more cases occurring in any one year is 0.501 (50%). The probability of 35 or more cases occurring in any one year is 0.046 (4.6%), whilst for 50 or more cases, the probability becomes 1.4×10^{-5} (0.001%).

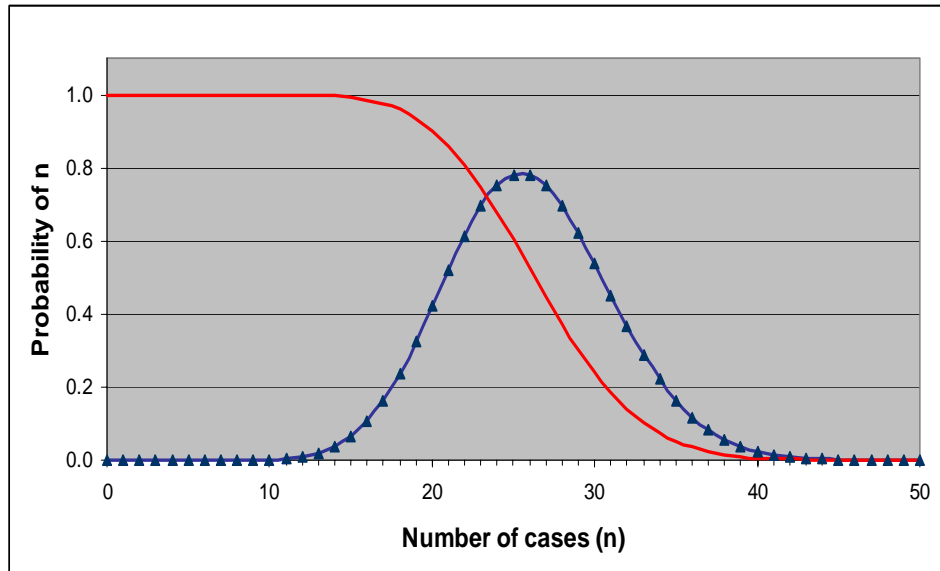
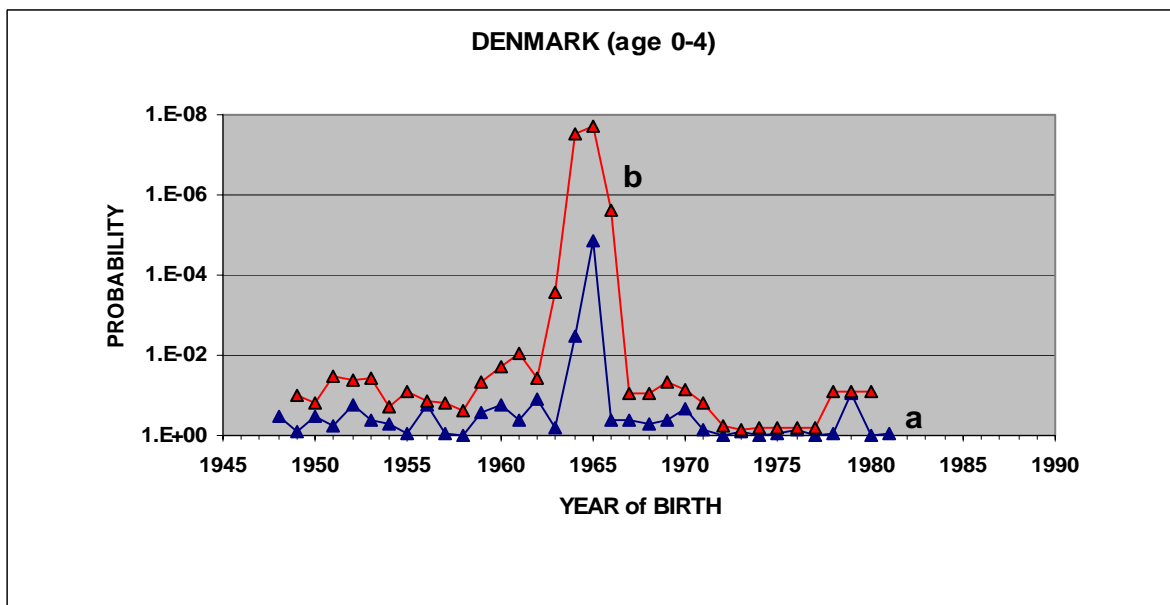


FIGURE 3. Probability of the observed number of cases arising in any particular year if these are random events following a Poisson distribution with average value 25.7. The lower (a) line gives the probability of single event, the upper (b) line the probability in year x of the three observed consecutive events in years (x-1), x, (x+1).



[‡] The Poisson distribution function is given by: $f_a(n) = \{a^n \cdot e^{-a}\} / n!$ where a is the average value of the variable n.

3. Analysis of Leukaemia Incidence by Children's Year of Birth

This analysis sets out to answer the question: was the incidence of leukaemia during the first five years of life abnormally high in any part of the study period in any of the Nordic countries?

For each of the remaining four Nordic countries, the data sets for children aged less than 5 (both sexes) for each annual birth cohort have been examined by the procedure are given in the preceding section.

1. For each country, the incidence of leukaemia in each annual birth cohort, up to the age of 5, is plotted by year of birth. Thus, line **a** (blue, usually upper line) shows the actual number of cases of leukaemia in children before their 5th birthday, and line **b** (red) in the same graph shows this incidence normalised to a population of 100,000.

An initial examination of the plots shows that the apparent anomaly in the Danish data, at around the year 1966, does not appear - at least in any marked a degree - in the incidence data for the other countries.

2. Examination of the probability plots (calculated as previously shown in detail for Denmark) again does not show any anomalies, either around the year 1965 or indeed at any point in the sequences.

Some further analysis (graphs) of the Danish data is given in Appendix II.

4. CONCLUSIONS

Statistical analysis of the data for leukaemia incidence in the Nordic countries, 1948-1984, in children under 5 shows an incidence rate around 7 cases per 100,000 population for each year-of-birth group (year cohort). For Norway, Sweden, Finland and Iceland, the annual variations lie reasonably within a Poisson distribution. However, there is an anomaly in the incidence data for Denmark, which is not reproduced in the analogous data from the other four Nordic countries.

The incidence of leukaemia in the 1965 and 1966 annual birth cohorts for Danish children under 5 shown an anomalously high incidence, which is so far removed from any likely random fluctuation as to suggest a specific cause. The fact that the anomaly is not replicated in the data for any of the other four countries could be taken to show that the "Danish anomaly" is unlikely to be a consequence of enhanced fallout, as this may be assumed to have fallen on all five countries. However, even this apparently obvious conclusion may be false if there are major differences in dietary habits. On the face of it, the anomaly must be related to some specific facet of the Danish environment and/or behaviour, and this conclusion would not exclude linkage to fallout.. No other potential causes appear yet to have been suggested.

I am most grateful to Professor Sarah Darby for supplying the data for this study, and to Dr Richard Wakeford for much helpful advice and discussion. However, I accept sole responsibility for any errors in the analysis and for the conclusions I have drawn.

JPD

28/11/03

FIGURES following:

Figures 4-7, incidence of leukaemia in children under 5 years of age (both sexes) for the annual birth cohorts for which data are available, for Norway, Sweden, Finland and Iceland, respectively (i.e. analogous to Figure 1, for Denmark).

Figures 8-11, probability plots analogous to Figure 3 (Denmark) for the four remaining Nordic counties.

Figure 4

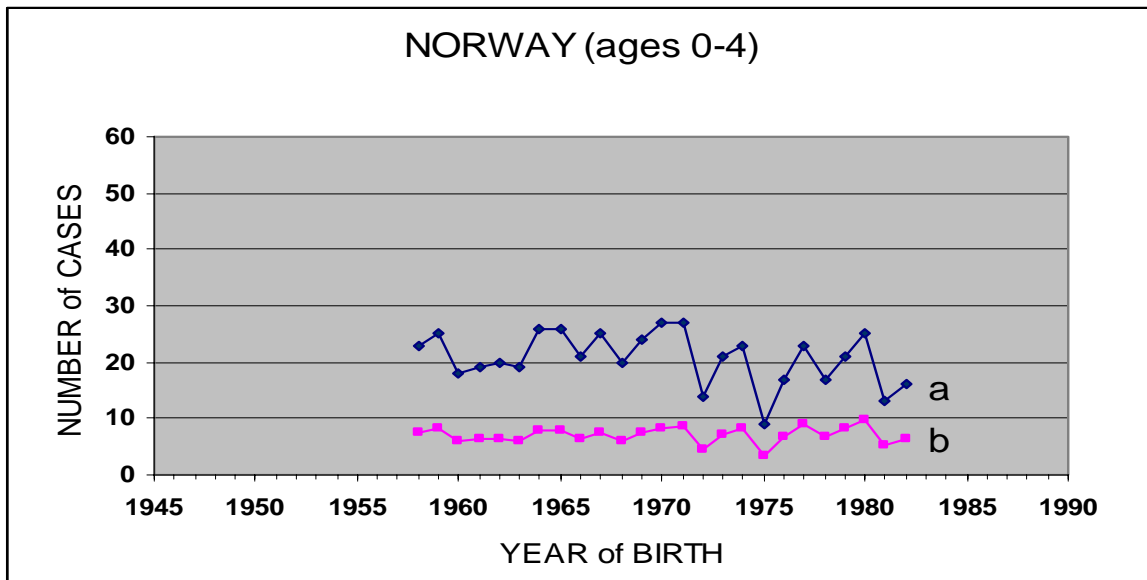


Figure 5

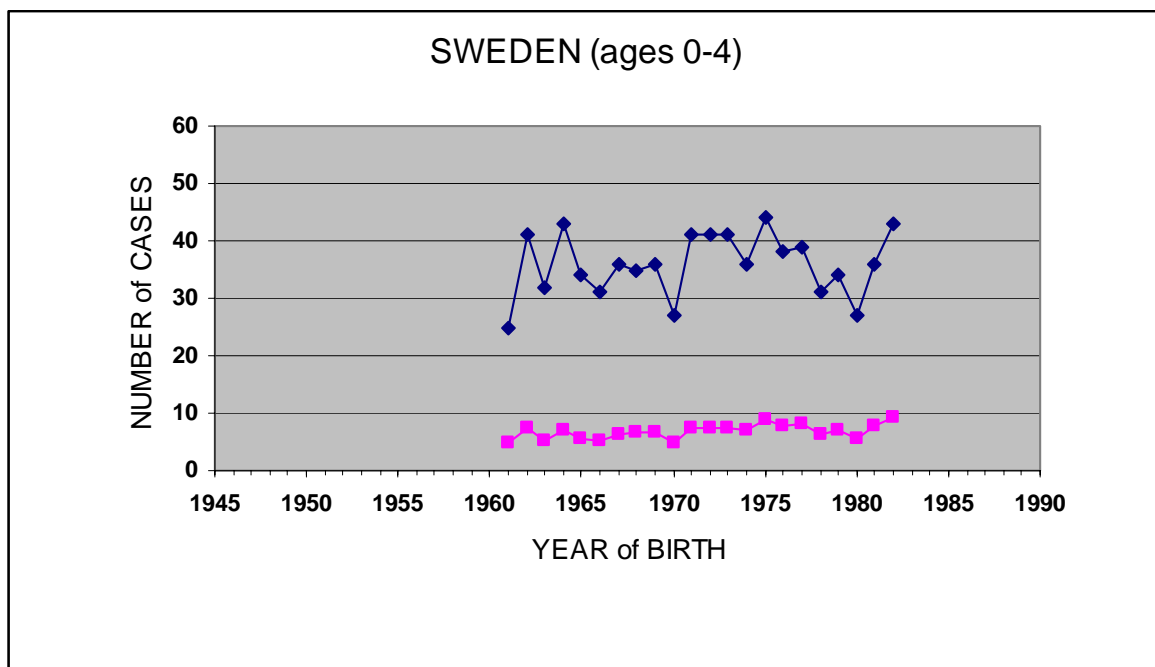


Figure 6

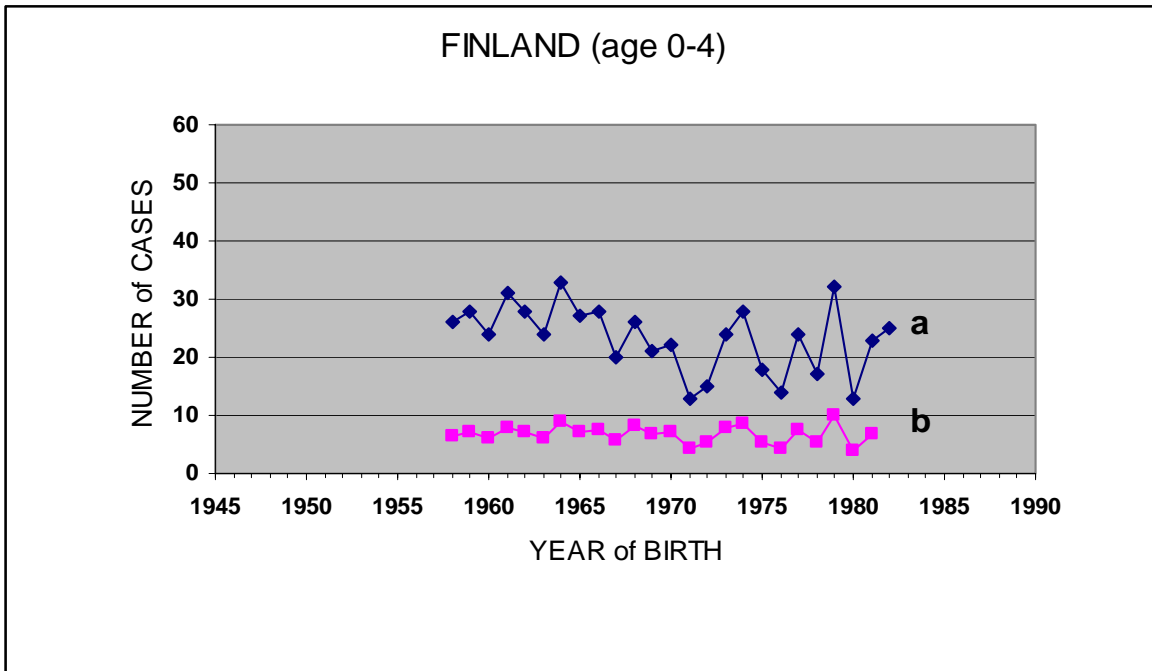


Figure 7

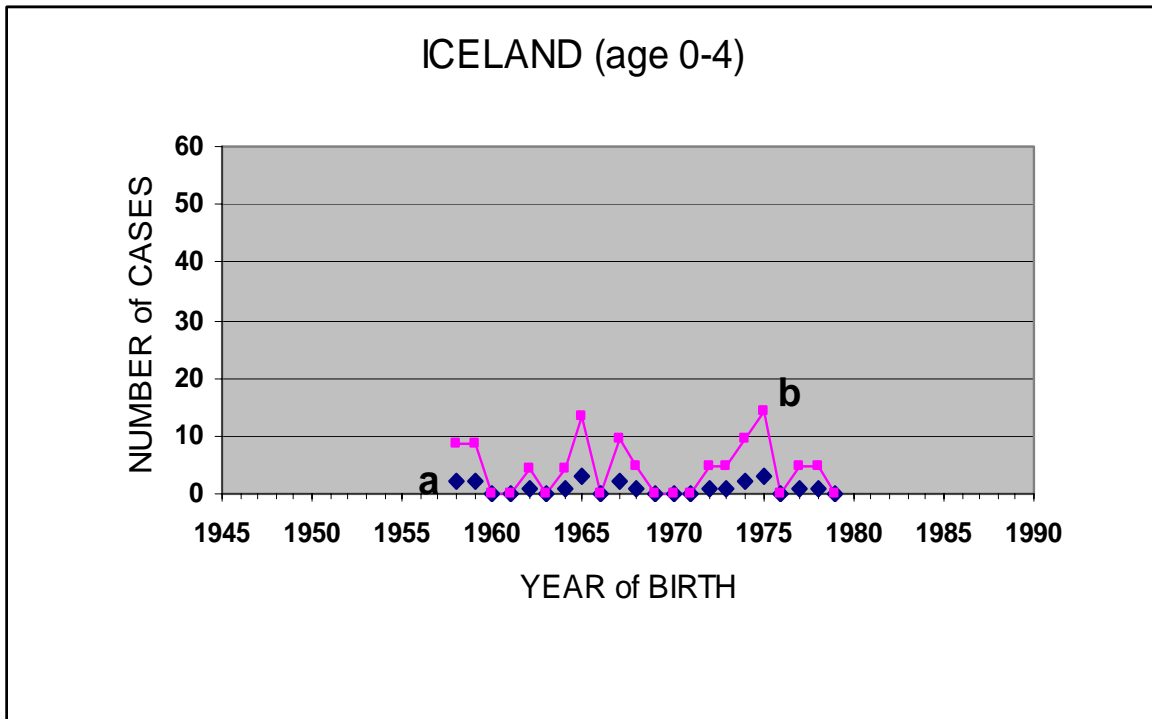


Figure 8

NORWAY (age 0-4)

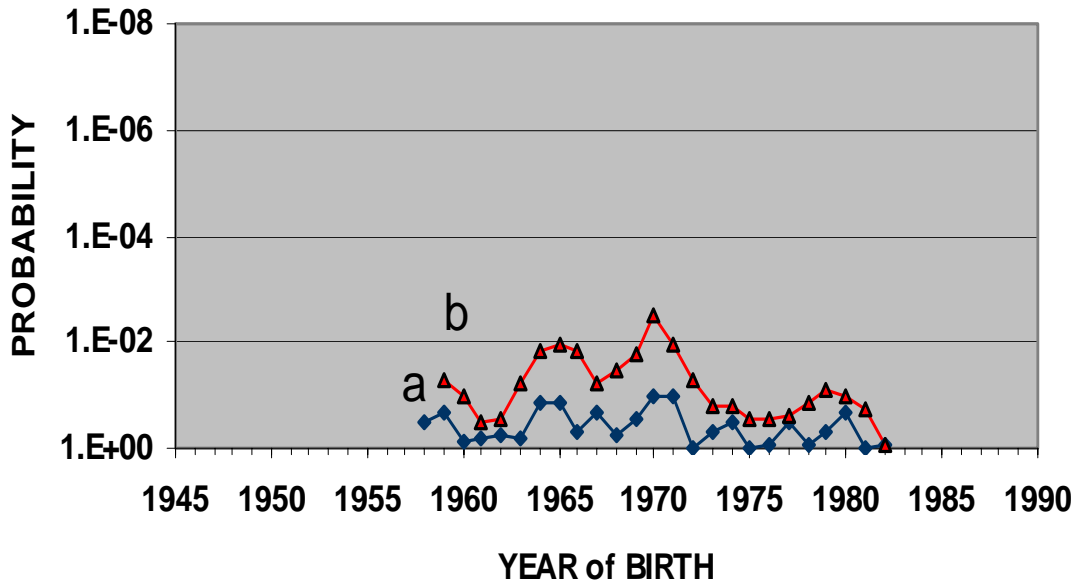


Figure 9

SWEDEN (age 0-4)

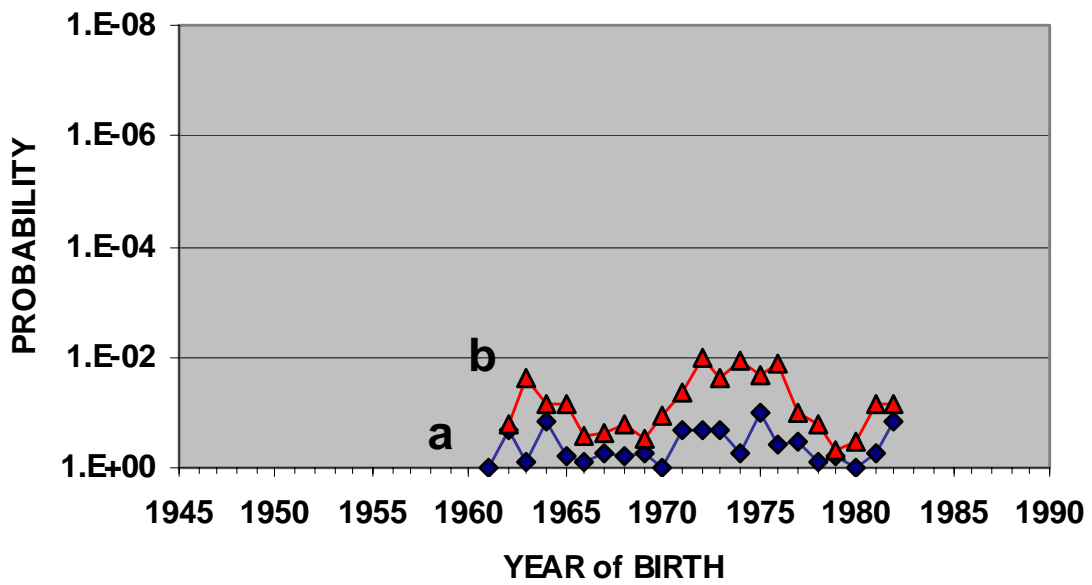


Figure 10

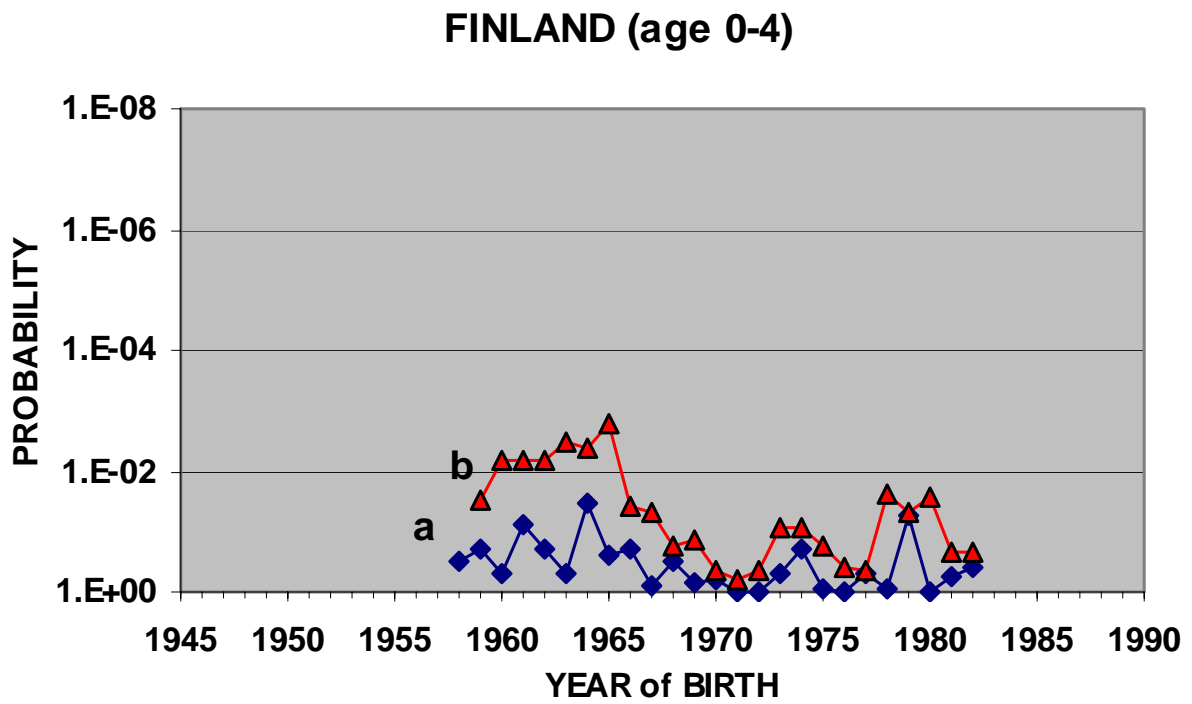
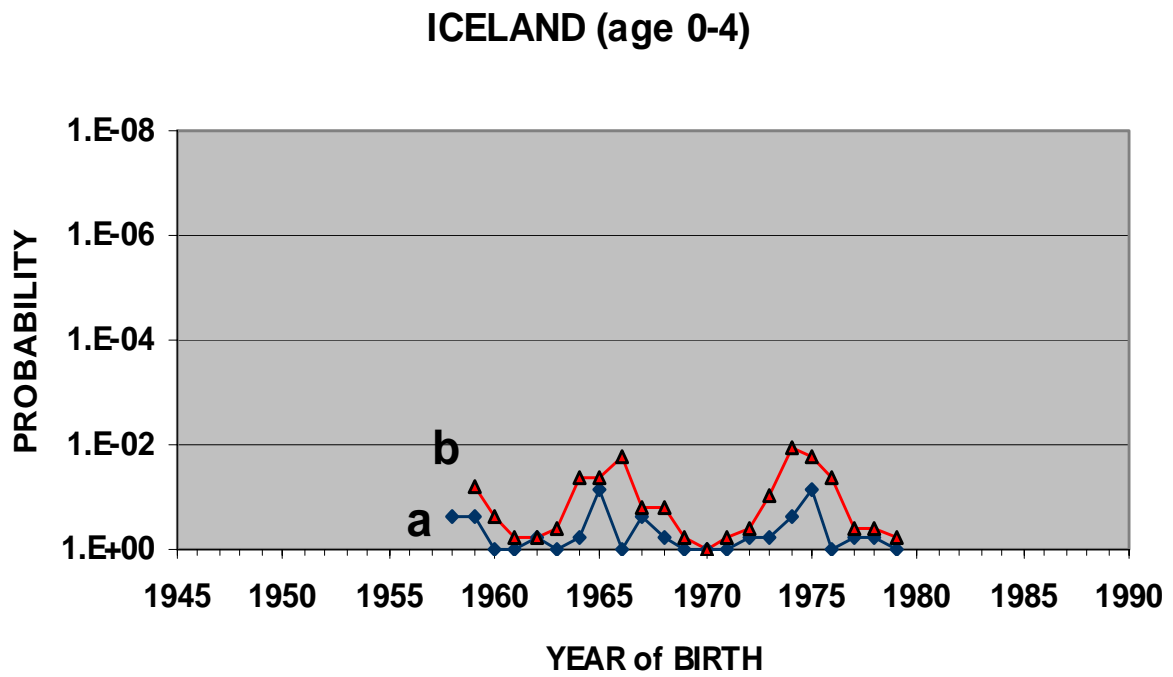


Figure 11

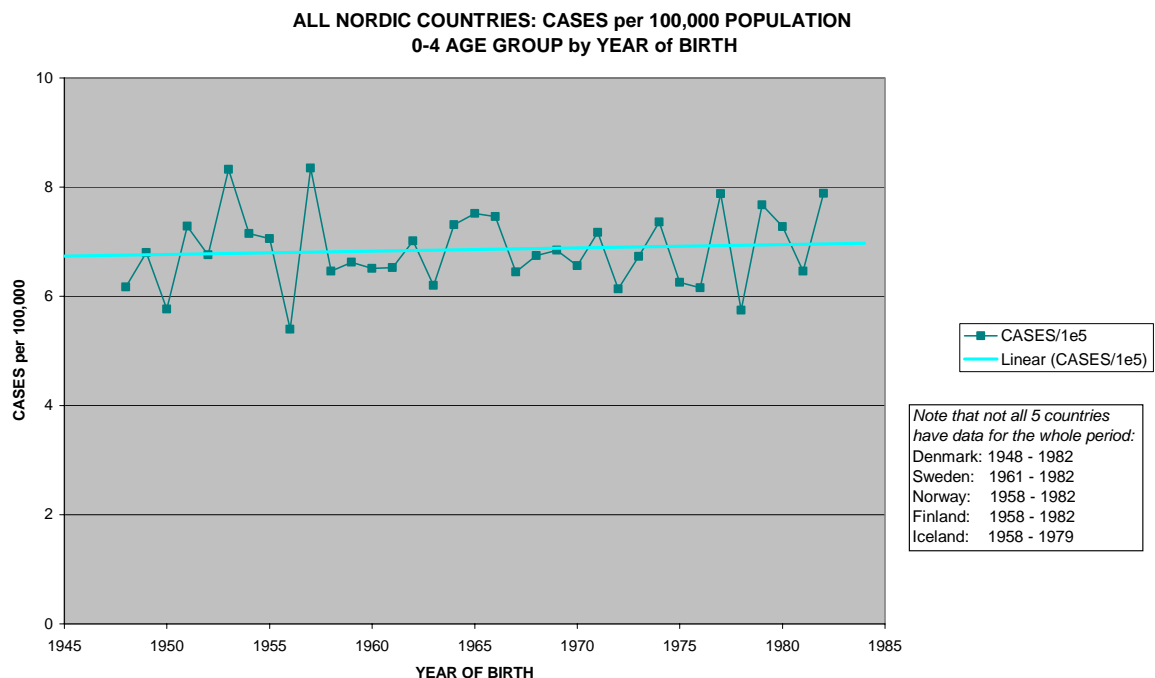


APPENDIX I

The following brief account is intended as a reminder to CERRIE of the basis for the discussions on this issue. The account reflects my understanding of the points made by the various parties, and if I have misrepresented any of these views, this was unintentional and I apologise. Please let me know if that is the case.

The paper by Darby et al. (BMJ, 1992) reporting on childhood leukaemia in Nordic countries showed, amongst other diagrams, a graph giving standardised rates of childhood leukaemia per 100,000 in the 0-4 age group (1948-82) for the 5 Nordic countries combined. A similar graph, derived from the data provided to me by Sarah Darby is reproduced below (Figure 12) and although not standardised (age, sex, country) is essentially the same as that shown in Darby et al., Figure 5.

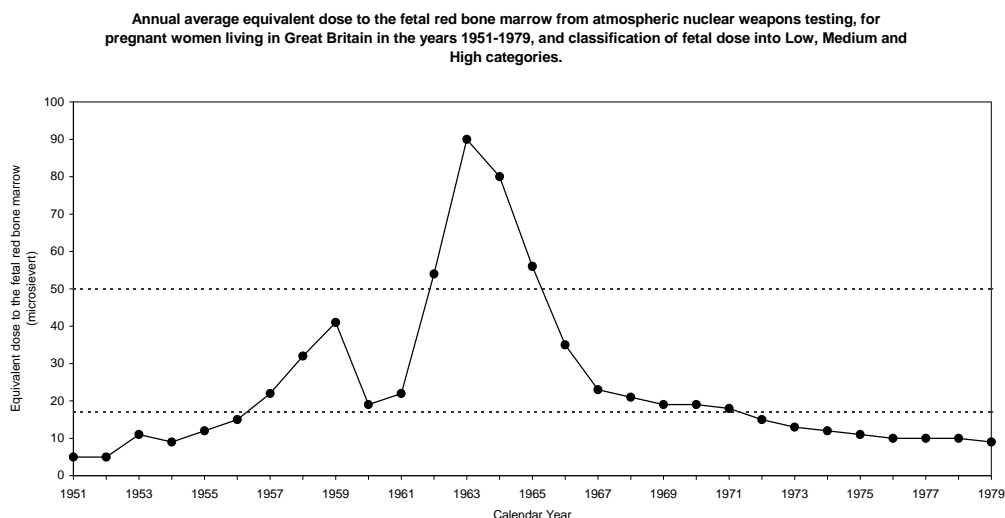
Figure 12. Leukaemia incidence by year of birth, all Nordic countries combined



Darby et al. concluded (amongst other things) that leukaemia incidence rates were slightly higher during the high exposure period, but that this was not related to estimated radiation dose to the red bone marrow (see for comparison Figure 13 for estimated doses from weapons test fallout in Great Britain).

The conclusions were challenged by Busby et al. (2003), who suggested that the combining of the data for all Nordic countries diluted a more pronounced effect shown by the Danish data, for which a far more significant rise in leukaemia incidence occurred over the high-fallout period.

Figure 13. Fallout doses in Great Britain over the period 1951-79 (graph supplied by Dr Wakeford).



The purpose of the present paper is simply to examine the data for leukaemia incidence for each of the five Nordic countries, to determine whether or not the data for any one country individually shows a significant rise in incidence rate at any period during the 1960s and 1970s. The test of "significance" is basically that of calculating the probability of any observed deviations from average in the annual incidence rate (i.e. the probability of deviations from the "horizontal straight line"). Clearly, because the numbers are small, deviations can be relatively large whilst remaining within the bounds of a reasonable probability. The analysis conducted assumes that the variations in annual incidence will (if random) follow Poisson statistics, and looks at the probability of any observed deviations on this basis, to determine whether or not the observed incidence rates in any of the Nordic countries show annual deviations which might be regarded as abnormal. The methodology is described in detail in the main text. There is no intention of attempting to infer any particular cause for any significant deviations which may be found to occur.

References

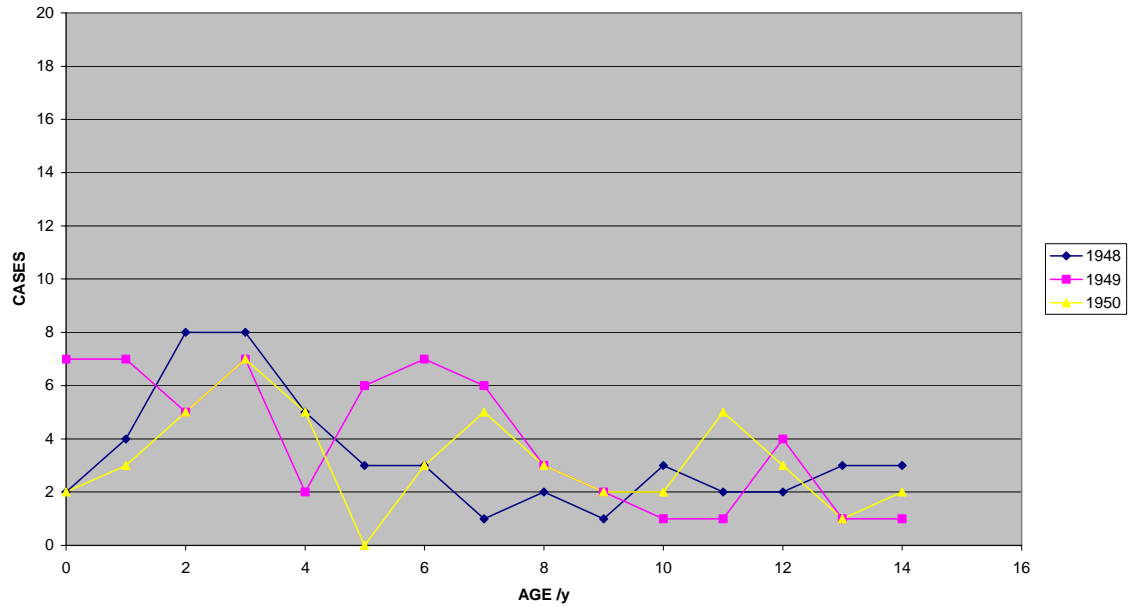
Darby SC et al., Trends in childhood leukaemia in Nordic countries in relation to fallout from atmospheric nuclear weapons testing. *British Medical Journal*, 1992, 304, 1005-1009

Busby C, Childhood leukaemia and global weapons-test fallout: the Nordic leukaemia study revisited. *Green Audit Occasional Paper 03/02*, Aberystwyth, January, 2003.

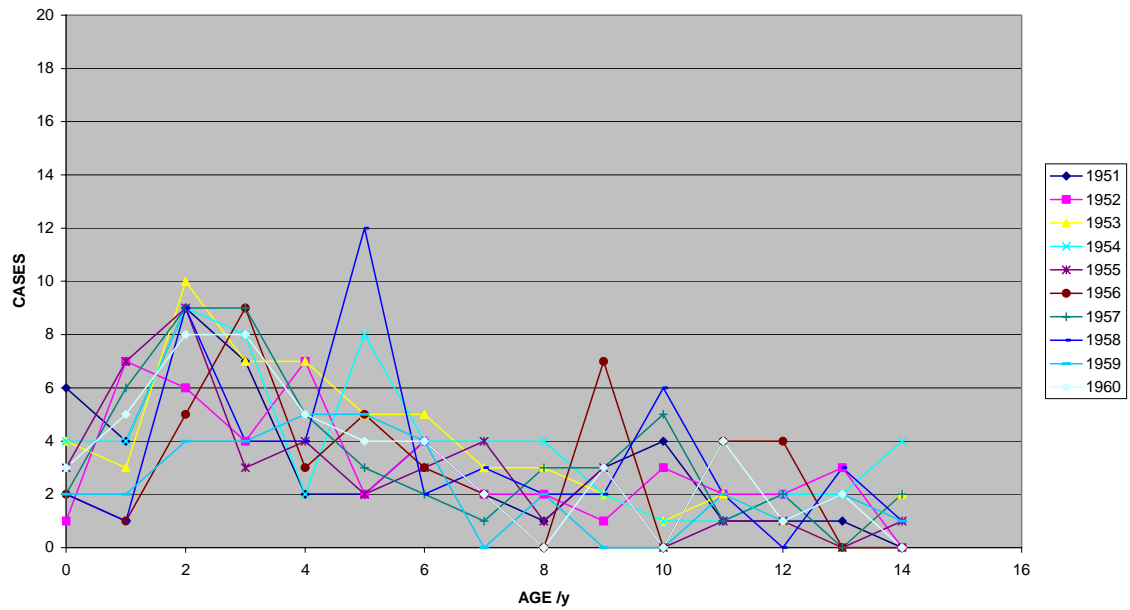
APPENDIX II

The following graphs show a more detailed analysis of the Danish data grouped by year of birth. Each plot shows the number of cases diagnosed against the age of diagnosis. The information is included for interest and is not further discussed.

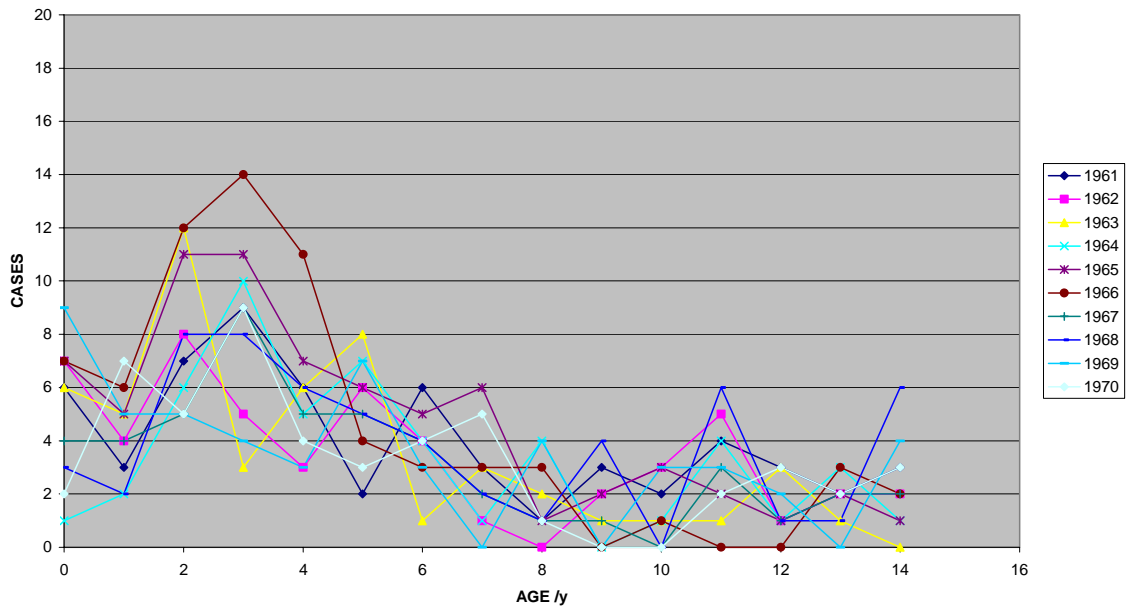
ANNUAL CASES by AGE



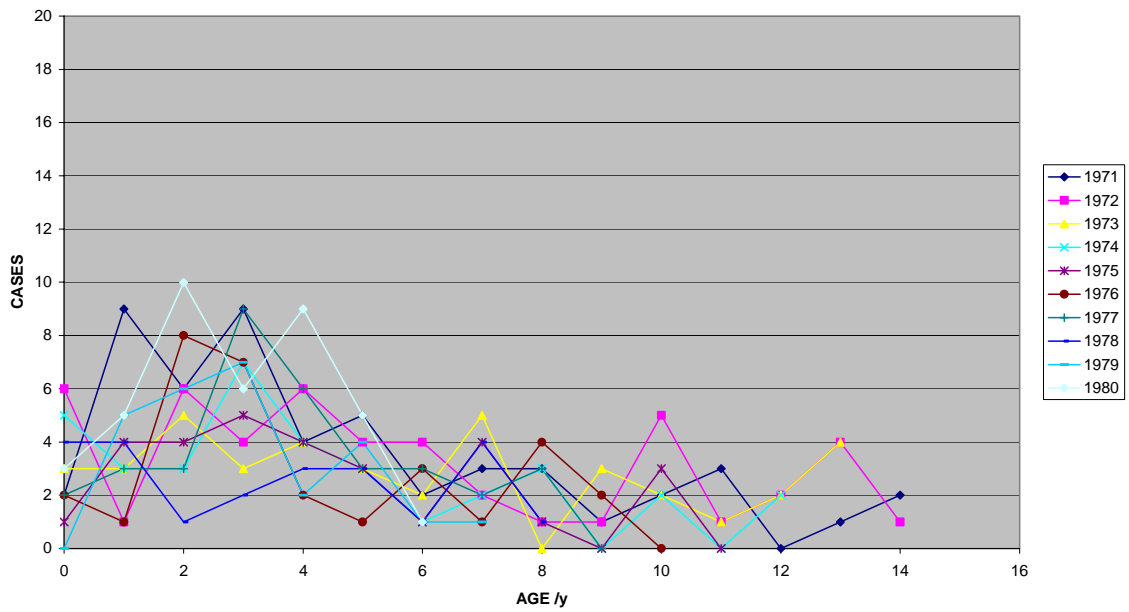
ANNUAL CASES by AGE



ANNUAL CASES by AGE



ANNUAL CASES by AGE



JPD
28/11/03