THE ROLE OF THE MEDICAL STATISTICIAN

Sally Hollis, Medical Statistician
Lancaster University

Medical statisticians may be employed to work on a specific medical research project, to work on a statistical project, to give general statistical advice to a wide range of people or a combination of these. This diversity of working situations is reflected in the range of interacting roles a medical statistician might have (Figure 1). Many benefits can be derived from these interactions. Statistical consulting work has a low status, and can seem like emergency first aid to revive dead projects which required resuscitation at birth, but it can also lead to more substantial collaboration with medical colleagues and their real problems and data, which in turn can inspire methodological work that is directly applicable to medical research problems.

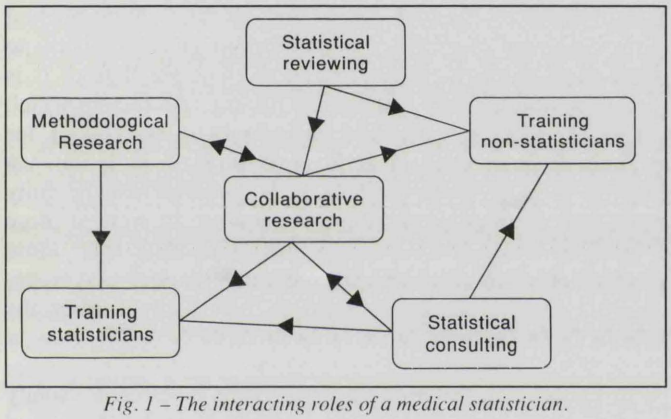

Fig. 1 – The interacting roles of a medical statistician.

An increasing number of medical journals are seeking statistical help and many now have a statistician on the editorial board. I am on the board of Diabetic Medicine and the Annals of Clinical Biochemistry: not an exciting job, but extremely necessary if high quality published clinical research is to be ensured. I think most other medical statisticians feel the same way, and since there is a shortage of suitable candidates for statistical editors, we are each obliged to do our bit! If nothing else, exposure to the problems in submitted work can be a helpful guide to the statistical training from which clinicians would benefit.

Statisticians and students benefit from exposure to a wide range of these types of experience. No one statistician can cover all of the above areas with any depth, so perhaps the ideal structure is a team of statisticians who collectively cover all of these areas and can each find their own balance. We hope that the Medical Statistics Unit at Lancaster will grow in this way – it has a good foundation with a very strong applied and theoretical research record (Box 1).

As part of the NHS Research and Development Initiative, the North West Regional Health Authority has helped to establish Medical Statistics Units in Lancaster, Liverpool and Manchester to support research in the region. The Lancaster Unit is a joint venture of Lancaster University and the Regional Health Authority and is based in the University's Department of Mathematics and Statistics. The Statistics group at Lancaster has a strong tradition of collaborative research in biomedical, environmental and social sciences, and was awarded the top rating in the latest research assessment of all UK Universities.

The Unit aims to:

- promote awareness of the statistical foundations of high quality health services research,
- train NHS professionals in the theory and application of statistical methods in health services research,
- provide statistical advice and expertise to NHS professionals through consultation and collaborative research,
- develop new statistical methodology relevant to health services research,
- train the next generation of medical statisticians.

Support given by the Unit may range from an answer to a specific question about some aspect of design or analysis to long-term involvement in a major project. The Health Authority are particularly keen to improve the quality of applications for research funding and we will be happy to advise on grant proposals.

I have been working in the Unit since its inception in April 1995 along with a part-time secretary (Cathy Thomson) and research associate (Barry Rowlingson). A Senior Lecturer (Dr Robin Henderson) will be joining us in October 1995. The Unit is directed by Professor Peter Diggle. If you would like further details or to make an appointment, please ring Cathy on 01524 594622.

Box 1 – Lancaster University Medical Statistics Unit.

HOW DO STATISTICS HELP THE MEDICAL COMMUNITY?

Childhood Leukaemia and Sellafield

To gain an insight into how statistical reasoning can guide the investigation of a medical question, it is instructive to consider the development of the evidence for a childhood leukaemia excess around Sellafield.

The debate was initiated by a Yorkshire Television programme 'Windscale: the nuclear laundry' in 1983. The producer had not set out to investigate this topic but heard of local suspicions about increased childhood leukaemia when he was researching effects of occupational exposure at Sellafield.

Yorkshire Television took care to verify the facts as far as they were able, and produced a "statistically significant" result. This was not, however, a planned study and the results must be interpreted with caution. The perceived increase may be due to a heightened level of local awareness of leukaemia brought about by one or two high profile cases. Seascale may even have a high incidence of childhood leukaemia due simply to natural geographical variation – some region in the country must have the highest rate even if there are no environmental factors at work.

In response to this allegation, the government set up an independent inquiry, chaired by Sir Douglas Black[1]. As well as examining currently available evidence, the inquiry was asked to suggest what further research was necessary. Martin

Gardner, a medical statistician at the MRC Environmental Epidemiology Unit at Southampton was asked to join the advisory committee and subsequently carried out studies recommended by the group[2].

The first step taken was validation of the cases using medical records not available to the television researchers. If some cases turned out not to be genuine then the allegation could be immediately refuted. Any statistical conclusions can only be as good as the data on which they are based. It was established that the submitted cases were real, and that no further cases existed.

The Black advisory group then examined studies of cancer incidence in Cumbria, including those in progress at the time of the report. These geographical studies showed an apparent excess incidence of mortality from childhood leukaemia in the Seascale area during 1968 to 1978 (Table 1). The expected number of deaths for an area at national rates could be calculated by multiplying the national death rate (deaths per thousand population) by the size of the area (population in thousands). Most death rates, however, vary considerably with age and sex. We can calculate a standardised expected number of deaths by carrying out the calculation separately for each age-sex group and summing the results. In order to compare the observed (actual) number of deaths with the expected number of deaths it is usual to calculate a Standardised Mortality Ratio (SMR), which is the observed number of deaths divided by the expected number of deaths. The result is usually multiplied by 100, so a number of deaths equal to that expected at national rates would give an SMR of 100 and a number of deaths twice that expected would give an SMR of 200. Crude death rates can be misleading; for example, Brighton has a high death rate but this is mainly due to the fact that retired couples often relocate to seaside resorts giving an older than average population. The standardised mortality ratio adjusts for this. The number of people in each age-sex group is usually estimated from census data.

Rural district*	Calendar period	No. of deaths		Standardised mortality ratio (SMR)	95% confidence interval for SMR
		Observed	Expected†		
Ennerdale	1959-67	3	3.3	91	19 to 266
	1968-78	4	3.3	121	33 to 310
Millom	1959-67	1	1.6	63	16 to 348
	1968-78	6	1.4	435	260 to 946

*Ennerdale contains the Sellafield nuclear site and Millom contains the village of Seascale.
†At age, sex and calendar period specific rates for England and Wales.

Copied from Gardner MJ. Investigating childhood leukaemia rates around the Sellafield nuclear plant. *Int Stat Rev* 1993;**61**:231-44.

Table 1 – Mortality from leukaemia under 25 years of age during 1959-78 in Millom and Ennerdale Rural Districts.

The apparent excess of mortality from childhood leukaemia in the Seascale area is difficult to interpret. Was this a fluke event or an exposure-related risk waiting to be discovered? Standard statistical procedures are designed to test *a priori* hypotheses rather than *post hoc* findings and the statistical P value cannot be interpreted in the usual way. One common statistical problem that arises particularly with *post hoc* hypotheses is that the excess risk and its statistical significance change with the selection of boundaries for geographical area, calendar years, age-groups and diagnostic categories. By 'moving the goal-posts' the results can be manipulated. Once we have observed the leukaemia cases it is difficult to define boundaries that are not influenced by the knowledge of the actual data. This is why it is good scientific practice to define the major hypotheses and how they will be tested **before** carrying out a study.

To appreciate the practical significance of the observed excess risk it is important to see the result in the context of other regions around the country. Leukaemia is a rare disease and there is bound to be considerable variation in the rates within small geographical regions. Comparison of the areas around Sellafield with the rest of the country showed that Millom was indeed atypical, the SMR of 435 being the second highest among 152 similar-sized rural districts around England and Wales.

The geographical data seem to indicate a raised childhood leukaemia rate near the Sellafield nuclear plant. The interpretation that radioactive discharges **caused** the increased risk is plausible since there is undeniable evidence that leukaemia can result from high exposure to radiation, but observational studies cannot establish a cause and effect relationship. A strong link would probably be observed between sunglass wearing and consumption of ice-cream but we would not infer that wearing sunglasses causes someone to eat ice-cream.

Seascale is not a typical West Cumbrian village. Much of the accommodation was built specifically to house staff of the nuclear industry. The population of Seascale is more mobile than that of many adjacent villages, which leads to large inaccuracy in estimates of the population size from the ten-yearly census data, affecting calculations of expected numbers of cases and SMRs. Cancers usually have a long latent interval between exposure to an environmental carcinogen and diagnosis. Population movement is likely to dilute any effects measured on the current population. It is clearly desirable to try and control for any risk factors which we know of or suspect. The major proven risk factor for childhood cancers is *in utero* exposure to radiation via medical X-rays. None of these considerations can be fully taken into account by observational studies. In view of these limitations, the Black inquiry recommended that more sophisticated epidemiological methods (cohort and case-control studies) should be used to investigate further the apparent excess in childhood leukaemia risk.

In a cohort study a group of individuals is selected, usually according to birth date and/or geographical location, and followed up over time. We can then compare the individuals who do and do not develop the disease in question. This has the advantage that the size of each group is known precisely. Cohort studies of all children born in Seascale or who attended schools in Seascale during 1950-1983 were carried out by Gardener et al[2]. Extensive searches were made of many different sources of information. All eligible children were followed up through the National Health Services Central Register. No personal contact was made with children or parents. Children in the birth cohort were excluded from the school cohort so that a comparison could be made of children born in the area and children who moved into the area after birth. Mortality and cancer registration in these cohorts were compared with that expected at national rates. There was a 10-fold excess of leukaemia in children born in the village but no cases in children who moved in after birth (Table 2). It appeared that there was a genuine excess of childhood leukaemia in Seascale and that if there is a local environmental factor, it must act very early in life or before birth[2].

Specific suggested risk factors can be examined using a case-control study in which a group of known cases is matched with a group who do not have the disease and their exposure to the putative risk factors is compared. This is often less demanding than a cohort study where a large number of individuals must be identified to yield the required

Diagnosis	Cohort	No. of deaths Observed	Expected*	Observed/ Expected	95% confidence interval for O/E
Leukaemia	Birth	6	0.6	10.0	3.7 to 21.8
	Schools	0	0.6	0	0 to 6.2
Other cancer	Birth	6	2.2	2.7	1.0 to 5.9
	Schools	4	3.4	1.2	0.3 to 3.0

*At age, sex and calendar period specific rates for England and Wales.

Copied from Gardner MJ. Investigating childhood leukaemia rates around the Sellafield nuclear plant. *Int Stat Rev* 1993;61:231-44.

Table 2 – Leukaemia and other cancer cases during 1950-86 in Seascale birth and school cohorts.

number of cases. Controls are usually chosen to match each case on age and sex and any other demographic variables which may be important. A case-control study was designed to examine established and potential risk factors for childhood leukaemia and exposure to radiation[2]. Eight controls were chosen for each case. Since the number of cases is fixed, the statistical efficiency of the study can only be increased by using more controls per case. Cases and controls were limited to births with a West Cumbrian address recorded for the mother. Two types of controls were chosen, firstly the four children of the same sex closest to the case in the birth register in each direction and secondly the closest four in each direction with the same maternal civil parish as the case. This allowed examination of geographical factors using the area controls and other factors using local controls which were more precisely matched.

The distance from Sellafield was calculated from the National Grid reference of each address at birth. Figure 2 shows the findings in sectors of increasing 5 km radius around Sellafield. Risks are given relative to the inner circle, so the relative risk of 0.17 in the outer sector indicates that children born in West Cumbria over 30 km from Sellafield have an estimated risk of developing leukaemia that is 17% of the risk in children born within 5 km of Sellafield. There was a large fall to about a third on moving outside the inner circle and some suggestion of a decreasing risk with further distance.

Information was collected about the cases and controls using medical records and questionnaires sent to the parents. For those parents who worked at Sellafield at any time, detailed information on radiation dose was supplied by British Nuclear Fuels plc. Analysis of fathers who had worked at Sellafield showed that the highest levels of paternal exposure were associated with about an 8-fold increase in risk of developing leukaemia (Figure 3). This is similar to the 10-fold increase found in the cohort study. We can see that the relative risk in the group with ≥100

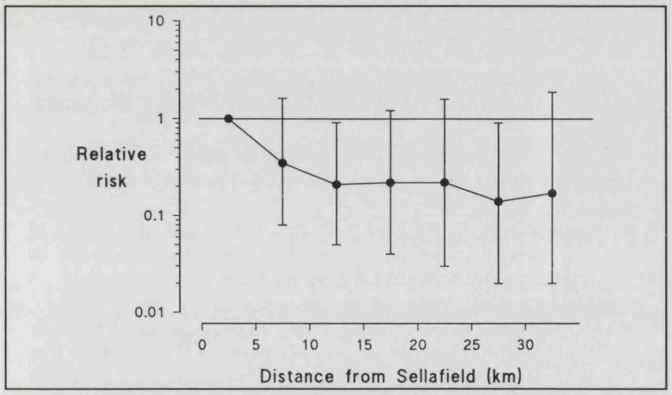

Fig. 2 – Relative risks for leukaemia in children by distance from Sellafield of residence at birth using area controls. (Drawn from data in Gardner MJ. Investigating childhood leukaemia rates around the Sellafield nuclear plant. Int Stat Rev *1993;**61**:231-44.)*

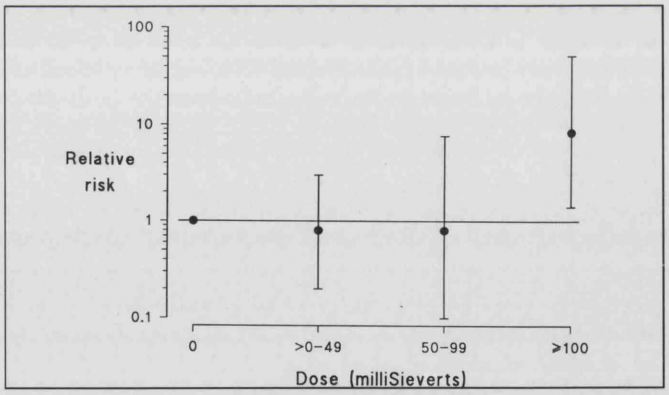

Fig. 3 – Relative risks for childhood leukaemia by father's external ionising radiation dose during employment at Sellafield before child's conception. (Drawn from data in Gardner MJ. Investigating childhood leukaemia rates around the Sellafield nuclear plant. Int Stat Rev *1993;**61**:231-44.)*

milliSieverts exposure is significantly different from the group with no exposure since its confidence interval does not include 1 which represents equivalence. Since this is a 95% confidence interval, we can conclude that $P<0.05$.

It was concluded that the father's exposure to radiation during his employment at Sellafield before the conception of his child may be relevant in some way to the subsequent development of leukaemia, and could explain statistically the excess cases occurring in the village of Seascale. Does this association indicate that there is a causal mechanism? This can only be answered with reference to other information on the effects of radiation. Some geneticists and radiobiologists have suggested that the levels of occupational exposure are too low for genetic damage to be a plausible mechanism on current knowledge. Some information can be obtained from studies of the children born to Japanese survivors of the atomic bombs, but these results may not be applicable since they relate to a high short-term exposure rather than a lower long-term exposure.

Another method which would shed more light on the causal nature of the relationship would be a randomised controlled trial. This would involve randomising a large number of men into two groups, one of which was then exposed to high doses of radiation. A subsequent excess of leukaemia in children of the men exposed to radiation compared to those not exposed would be much more convincing evidence of a causal link since through randomisation we have ensured that the two groups are similar in other respects. However this would obviously not be possible due to ethical and practical considerations.

Geographical studies around nuclear establishments in other countries have not generally shown local excesses or relationships with the father's exposure. In general we look for consistency of findings across space and time as evidence of a "real" difference, but we are not dealing with a well-defined phenomenon and we cannot be sure that Sellafield does not have some unique element not found in other nuclear establishments.

In 1992, two legal cases were brought against British Nuclear Fuels, the plaintiffs claiming that exposure to ionising radiation resulting from operations at Sellafield was the cause or a material contributory cause of a fatal infant leukaemia and a non-fatal non-Hodgkin's lymphoma. The cases were considered concurrently and evidence was heard from many expert witnesses. The epidemiological and genetic evidence was considered of particular importance. There was discussion about how a novel statistical association should be assessed scientifically and whether a causal interpretation could be sustained in this instance[3]. The

current evidence was examined very closely including further follow-up of control subjects who had not been traced in the original study and re-analysis of the data by the defendants' expert witnesses. Many points specific to the two individual cases were discussed, but also of crucial importance was the hypothesis that childhood leukaemia arises as the result of the action of an infectious agent and that in communities where unusual population mixing has occurred, leading to unusual infective patterns, the risk of childhood leukaemia may be raised[4]. In the end, the judge found that the evidence was 'decisively' against preconceptional irradiation being a material contributory cause of either the two specific cases or the Seascale excess.

The questions of the existence and cause of an excess in childhood leukaemia around Sellafield were clearly not simple. Careful data collection, study design, statistical analysis and interpretation are all necessary to find unbiased answers in a situation where we cannot carry out an "experiment". Medical statisticians have experience of these problems and can help clinicians to answer difficult questions.

HOW CAN A MEDICAL STATISTICIAN HELP YOU?

A medical statistician can provide help with many stages of a research project:

- Study design
- Efficient analysis
- Presentation of results
- Interpretation of findings

Study design is extremely important, since no amount of sophisticated statistical analysis later can compensate for major design flaws. Designing a study entails specification of the study sample, size of sample, method of treatment, allocation to treatment (randomised, cross-over, etc) and choice of outcome measures. Careful design is particularly vital for observational studies. It is important that the sample is representative of the population or the results will be unreliable. The infamous Lanarkshire milk study[5] illustrates some of the biases that can occur. The study set out to determine whether provision of daily milk would affect the height and weight of school children. From February to June 1930, 10,000 children received 3/4 pint of milk daily while another 10,000 were selected as controls. The children were weighed and their height was measured at the beginning and end of the experiment. The children getting extra milk gained significantly more weight than did the controls. However, on closer examination of the results "Student" (of Student's t-test) pointed out several problems which could mean that the difference in weight gain between the groups was not due to the extra milk alone[6]. The controls were taller and heavier at the beginning of the experiment than those getting extra milk. The teachers had been allowed to adjust the allocation to "obtain a more level selection" and it appears that they enabled the poorer children to get the extra milk. The children all seemed to grow less than would be expected in a four month period. This is probably due to the simple fact that most children will wear lighter clothing in June than in February. Since poorer children are likely to wear relatively fewer clothes in winter this may have a complex effect on the measurements. Errors like these are very easy to make and it is always worth consulting a medical statistician before embarking on a study even if it is not quite as ambitious as the measurement of thousands of school children.

Good presentation of data can be extremely illuminating, both to identify interesting or unusual features during analysis of the data and to present powerfully the message of the results to a reader. Figure 4 shows scatter plots of four artificial datasets[7].

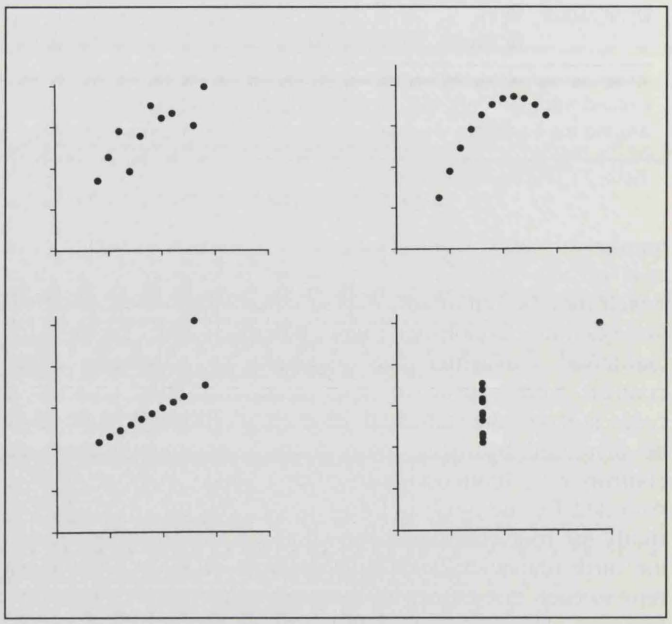

Fig. 4 – Four artificially-constructed datasets[7]. Which shows the strongest statistical relationship?

- Which do you think shows the strongest statistical relationship?
- What conclusions do you think you might have drawn if you knew the mean and standard deviation of the x and y variables, the line of best fit and the correlation coefficient but you had not seen the data graphically?

Now read the footnote at the end of the article.

If you are a "user" of research rather than a "doer", statistics is still important. You will need to understand and interpret the mysteries of $P<0.05$ and "power". I find it extremely helpful to use the analogy of a legal trial to illustrate the concepts of significance testing in an informative and memorable way (Box 2).

Innocent until proven guilty:
We always start by assuming that the null hypothesis of no difference (H_0) is true.

Beyond reasonable doubt (Type I error):
P is the probability of observing the given data if H_0 is true. We say $p<0.05$ is significant = beyond reasonable doubt. We have set our level of reasonable doubt so that we will "wrongly convict" one in twenty cases.

Insufficient evidence (Type II error):
The Director of Public Prosecutions often rules out prosecution due to insufficient evidence. This is because the chance of getting a conviction will be small even if the person is guilty.

Power is the chance of detecting a given effect size in a study. Studies should not be carried out without sufficient power.

Judge decides sentence:
Although the jury decides the verdict, it is left to a judge to allocate the appropriate punishment. The consequences of significance tests should be judged clinically.

Biases in evidence make verdict unreliable:
The conviction of the Guildford Four relied heavily on forensic evidence of explosives on the suspects' hands. When this evidence later proved to be incorrect, the original guilty verdict was questionable.

Box 2 – Significance testing as a legal trial.

Sir Austin Bradford-Hill[8] gives a very good example of how statistical results need careful interpretation and not just a test of significance. Figure 5 shows possible results of trials to detect whether each of five specific drugs reduces blood pressure. Let us suppose that you are only interested in prescribing drugs which reduce blood pressure by at least 5 mm on average. A ninety-five percent confidence interval for the mean change in blood pressure is shown for each trial.

- Which drugs give statistically significant results?
- Which drugs give clinically important results?
- Which drugs are worth investigating further?
- Does each of these questions suggest the same drugs?

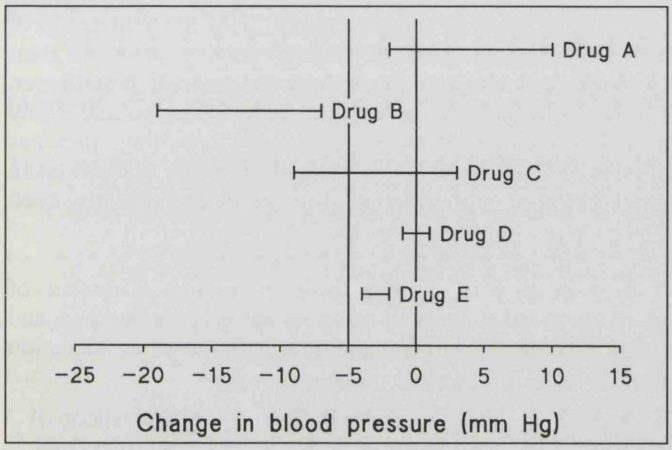

Fig. 5 – Hypothetical change in blood pressure from five drug trials (95% confidence limits)[8].

See Box 3 for some thoughts on how you would answer these questions.

Effects are statistically significant if the confidence interval does not include zero:
Drugs B and E give statistically significant results.

Effects are clinically important if the confidence interval lies entirely in the clinically important area – in this case to the left of –5mm:
Drug B gives a clinically important result.

Effects are worth further investigation if the confidence interval includes a clinically important effect – in this case a reduction of over 5 mm:
Drug C is worth further investigation with a larger sample size.

Statistical significance and clinical importance are not equivalent:
The effect of drug E is statistically significant but not clinically important. The effect of drug C may be clinically interesting but is not statistically significant.

Box 3 – Statistical significance and clinical importance.

I hope I have convinced you that medical statisticians can help not only with difficult problems of complex medical research but also with presentation of simple data and even with the interpretation of other people's research.

REFERENCES

1. Gardener MJ. Investigating childhood leukaemia rates around the Sellafield nuclear plant. Int Stat Rev 1993;61:231-44.

2. Black D. Investigation of the possible increased incidence of cancer in West Cumbria. London: HMSO.

3. Wakeford R, Tawn EJ. Childhood leukaemia and Sellafield: the legal cases. J Radiol Prot 1994;14:293-316.

4. Kinlen LJ, Hudson CM, Stiller CA. Contacts between adults as evidence of an infective origin of childhood leukaemia: an explanation for the excess near nuclear establishments in West Berkshire? Br J Cancer 1991;64;549-54.

5. Leighton G, McKinlay PL. Milk consumption and the growth of schoolchildren. Edinburgh: HMSO.

6. "Student." The Lanarkshire milk experiment. Biometrika 1931;23:398-406.

7. Anscombe FJ. Graphs in statistical analysis. Am Stat. 1973;27:17-21

8. Bradford-Hill A, Hill ID. Principles of Medical Statistics. London: Edward Arnold.

All four datasets have a correlation coefficient of 0.82. The regression lines and the means and SDs for the x and y variables are identical!