

STATISTICAL SIGNIFICANCE TESTING AND THE BLACK REPORT ON SELLAFIELD

History

On 1st November 1983 YTV broadcast a programme called "Windscale - the Nuclear Laundry" in which they examined the incidence of radiation-related diseases and abnormalities in the area surrounding the British Nuclear Fuels Ltd (BNFL) plant at Sellafield (for merly Windscale) in Cumbria. The programme showed much material that must have been deeply worrying to anyone, and especially to parents of young children, living in the neighbourhood of the plant. In particular it claimed that there was an incidence of childhood leukaemia cases of the order of 10 times that which would be expected purely by chance: in a discussion afterwards, one of the discussants, described as a 'medical statistician' claimed this represented a 1 in 10,000 chance.

On the other hand, in an article published in the Lancet, Craft and his co-workers (1) tended to dismiss such concern, claiming that it was of little value to consider the incidence rates for small areas.

The government, naturally disturbed by the allegations in the programme, reacted quickly by setting up an Independent Advisory Group under the Chairmanship of Professor Sir Douglas Black which collected data, held deliberations, reported and published by July 1984. Their report (2), perhaps predictably, took a middle path between the extreme positions of YTV and that of Craft and co-workers, offering 'qualified reassurance'. Controversy has continued since, particularly in the Lancet and New Scientist (3) This article is an attempt to assess, in a relatively non-statistical way, some of the statistical issues which have arisen.

The Report, which I shall refer to as the Black Report, may be divided very roughly into a radiobiological section, which attempts to trace and quantify the effects and the possible causal pathways whereby radiation from the BNFL plant could affect the health of the surrounding populations, and an epidemiological section which endeavours to make an

assessment of whether in fact there is a significant excess of cases of childhood leukaemia in the area.

The first of these sections of little value since it comes to the conclusion that the excess of cases is some 40 times that which would be predicted by their model and this can be safely ignored. Common sense should have told them that this represents gaps in their model, and is thus an even more damning indictment of the effects of the plant. Taken to extreme, the argument of the first section would imply that had we found ten times as many leukaemia deaths, we could have been quite confident that the plant was safe.

Statistical issues discussed by the Black Report

The epidemiological section on the other hand is much better thought out and clearly explains some of the difficulties in interpreting such data. Apparent epidemics can in some instances be artefacts of data collection or presentation. Rare events have to occur sometime and somewhere; people are more likely to be aware of and thus diagnose diseases which are in the news; it is a natural human quality to stretch a regional or temporal boundary to include a case occurring in the next village, or the year before the time period which it was planned to investigate. The Report explains these potential complications clearly in layman's terms. What its authors fail to do, (and this is a serious omission), is to discuss satisfactorily the statistical basis for the application of significance testing to such a question.

The Data

The Black Report lists

' 32 cases of leukaemia and other forms of cancer known to us by 1st June 1984 in the under 25 year old population of Millom Rural District based on information from YTV, death certificate data and the preliminary findings from an intensive review of hospital records being carried out for us by the West Cumbria Health Authorities'. (4)

These are displayed in Tables 2.1 - 2.4 of the Report. It is not always clear how these are related to the analyses, and indeed the analyses have been criticised as excluding some cases, (5) so it will be useful to examine how some of the numbers arose. I concentrate on two analyses which appear to have been particularly influential, those of Craft and Openshaw, and Gardner and Winter.

Craft and Openshaw (6) investigated the incidence of all cancers, and in a separate analysis all lymphoid malignancies (7) in the Northern Children's Cancer Registry Area (8) for children aged 0 - 14 in the period 1968-82. Incidence in Seascale was compared with that in the other 765 electoral wards in the area. The top ten are shown in Tables 2.18 (all cancers) and Table 2.19 (lymphoid malignancies). The year 1968 seems to have been chosen as the start as it was the date of the start of the Northern Children's Cancer Registry and 1982 as the end as this was the last year for which complete data was available (9) Table 2.19 shows that the incidence of lymphoid malignancies was 4. The breakdown of the Black cases seems to be as follows; of the 32 of childhood cancer cases listed by Black, 14 were classified as leukaemia and of these 7 were classified as being resident in Seascale at the time of diagnosis. Of these 2 were diagnosed before 1968 and one was aged 20 at diagnosis, leaving 4. The same 4 cases formed 'all cancers' incidence, one further case being excluded since it was diagnosed after the fifteenth birthday.

Gardner and Winter examined for all ages, and for 0 - 24; death from all causes, from all cancers (including leukaemia) and from leukaemia (not including lymphomas), in Ennerdale RD (containing the Sellafield site) and Millom RD (containing Seascale village) for the periods 1957-67 and 1968-78 and compared these with other areas in Great Britain. The data that they use was obtained from OPCS for 1959-79 for the Atlas of Cancer Mortality.(10) The 1968-78 period corresponds with the date of the results shown in that publication.

Apparently the 32 Black cases were classified thus; 17 were excluded because they died outside the study period, or were still alive and two because information was not available on date or place of death.

This left 13 cases of which one was leukaemia and 2 were other cancers in 1959-67, and 6 were leukaemia and 4 were other cancers in 1968-78. Two observations can be made here. Firstly the cases actually included in the analysis are very many fewer than those originally submitted to the committee. On the disputed assumption that the data is correct this is simply because careful analysis necessitates restricting the time period and area to allow comparability with other areas.

Secondly the finding that the data collected by the Children's Registry or by OPCS in the standard way can be reconciled with that collected from all possible sources for the Black committee is encouraging for the validity of the comparable data for areas other than Seascale.

The Findings

Confining attention to the area around Sellafield, the Report described the incidence of all cancers, and, within these, leukaemia, and the incidence for all ages and for the 0 - 24 age group from Millom, the Rural District containing Seascale for 1959-67 and 1968-78 (Gardner and Winter) and for the 0-15 age for Seascale, the five coastal districts and the rest of Millom RD (Craft and Openshaw). Their results are summarised as suggesting;

- (a) 'an approximately four-fold (higher) rate of leukaemia mortality in the under 25 year old population in Millom Rural District during 1968-78 or two-fold during 1959-78 (Gardner and Winter, 1984) - and an approximately ten-fold higher rate of leukaemia incidence in the under 10 year old population of Seascale' (Para. 2.33) and that

(b) 'for the years studied Seascale ranked sixth in incidence rates for all childhood cancers and had the third highest rate of childhood "lymphoid malignancies" among the 765 electoral wards, (in the Northern Childrens Cancer Registry Area)' (Para. 2.37). The apparent effects seem to be concentrated among children as the 'all age' effects are not statistically significant for these areas.

No unusual effects were noted for Ennerdale, the RD actually containing Sellafield and the area to the North and East of it.

The results are difficult to interpret because of the small number of cases: for example there are only four childhood leukaemias in 1968-82 in Seascale.

Urquhart and Cutler (11) allege that some of these exclusions were unjustifiable and that the Black results are thus underestimated, and unnecessarily reassuring. They claim that of the 17 cases on the Windscale coast submitted by Yorkshire TV to the Black Inquiry, 7 were wrongly omitted from the calculation of cancer rates, including 2 cases of leukaemia in Seascale. The allegations were discussed by Craft and Openshaw and Gardner (12). Of the seven 'disappearing' cases fuller explanations show that one is straightforward, but the others raise some interesting questions.

(a) One case was in fact included in the analyses.

(b) Two cases (of leukaemia in Seascale) were diagnosed in 1983 and 1984, respectively. Since the Committee did not have at its disposal comparable figures for the other areas, it would have been contrary to usual statistical practice to include them as this would have meant inflating the numbers of cases in Seascale

without comparable treatment of the other areas. Two cases of leukaemia in two years in an area with only about 400 children is a serious incidence, and should not be simply dismissed on these grounds without considerable thought. I shall return to this question in a later section.

(c) One case was wrongly coded in the Craft & Openshaw study as outside Cumbria (correct coding would not have situated it in Seascale but in Bootle, on the Cumbrian Coast). Given the urgency with which the analyses were required, and the consequent difficulty in carrying out thorough checking, this error was unfortunate but understandable.

(d) Two cases were recorded in the Cancer Registry as having been diagnosed after their fifteenth birthday, contrary to the ITV evidence. Careful checking of this anomaly is required: either source could be in error.

(e) One case was diagnosed as erythroleukaemia rather than leukaemia, Cutler argued that it should be classified as leukaemia. Gardner (13) explains that this case was omitted from his analysis since at the time of the diagnosis, erythroleukaemia was not classified as leukaemia on death certificates, and the rates for comparison areas would exclude leukaemia.

Thus with the exception, of course, of (c), the exclusions are justifiable on the grounds of common epidemiological practice.

Assessment of the controversy

In another context (Reading Between the Numbers - Radical Statistics Education Group 1982) I was a member of a team which attempted to draw up a checklist of some guidelines for assessing education research. It is of some interest to apply them to the Black Report as it is equally controversial and affects the lives and livelihoods of a large number of people.

We listed eight 'statistical responsibility' criteria which would be used to help assess whether an education study merits attention. These are:-

1. construct validity
 2. statistical conclusion validity
 3. internal validity and 'third variables'
 4. external validity
 5. level of analysis
 6. explanations of statistical techniques appropriate for the intended audience
 7. discrepancies in the care given to interpretation in different parts of the report
 8. fabrication of results and deception in presentation
- Also, to emphasise the fact that research is a social product with social impacts we added a further set of five questions which focus on the formative stages of the research and the ways in which it is put to use in the society. These are:-

1. Who commissioned and sponsored the research?
2. How was the research problem conceptualised and how well did the conceptualisation and the measures used in the study fit together?
3. How was the research presented and received?
4. How was the research used?
5. Are there alternative uses of the research?

I concentrate on the statistical parts of the Report. The comments are only my own opinions: reader may disagree.

- 1) Construct Validity. Are the measures or indicators chosen to represent the underlying concepts of the research appropriate? The study aimed at;
 - 'a. establishing the incidence of cancer in the area adjacent to Sellafield, and comparing it with the incidence of cancer in other areas in the United Kingdom and in Cumbria;

- b. considering the available data on radiation exposure in the area adjacent to Sellafield and the evidence relating to radiation exposure to cancer, thus assessing the likelihood that any radiation exposure could have caused any increased incidence of cancer detected in the area;
- c. assessing other possible significant factors'. (Black, Para. 1.6)

The definition of cancer was a standard grouping of the ICD (International Classification of Diseases). As discussed earlier, there was a query about whether one further case should have been classified as leukaemia but insofar as there was any doubt, omission was justified on the important grounds of comparability with other areas, since the comparison with other areas was a main feature of the study.

In the statistical section of the Report, the measure of exposure has effectively been taken as number of children in the area. A more sophisticated analysis would have involved the consideration of 'child-years' of exposure. Similarly deaths due to Sellafield pollution were taken as deaths in the area, whereas one ought to take into account any cases of children who left the area and subsequently contracted the disease (or died of it, depending on the outcome measure). The writers of the Report are of course aware of weaknesses. (See Recommendations, page 93).

- 2) Statistical conclusion validity

The validity of the statistical conclusions can be challenged on two grounds

- (a) the ranking of the results
- (b) the grounds for the type of significance test used.

(a) Ranking of the results

Urquhart & Cutler (14) have queried the method of ranking of electoral wards in which the top ten were shown in Table 2.19 of the Report, arguing that the ranking should have been in order of Poisson probability (i.e. significance), rather than in order of incidence rate as actually reported. This would have the effect of ranking Seascale as first rather than as third as in the Black Report. Their confusion is understandable, given that Craft and co-workers elsewhere reported the top ten of some results ordered by Poisson probabilities (15) However, the fundamental quantity of incidence is the rate, and the significance, which is a blend of observed magnitude of effect and size of group on which it was observed, is more an indication of how much reliance may be placed upon its magnitude. A more serious question relating to significance testing is the underlying philosophy of the type of test chosen. To discuss this it is necessary to review the procedures available when a number of possible quantities are to be tested.

(b) A priori and a posteriori comparisons

Statistical significance testing sets up a null hypothesis of no real effect and then investigates how often an apparent effect could arise solely as a result of sampling fluctuations. If this is less than some pre-determined level, normally 1 in 20, 1 in 100 or 1 in 1000, then the effect is deemed to be unlikely under the null hypothesis and thus the null hypothesis itself is hence deemed unlikely. Obviously if one carried out enough repetitions of comparable comparisons, sooner or later one would pass the magic significance level. Thus when Craft and Openshaw in the Black Report looked at 765 communities it was not surprising that at least one of them had statistically significant results at the 1 in 1000 (P < .001) level.

When one browses through a number of elements and picks out the largest one(s) it is necessary, to avoid cheating, to use the relatively weak a posteriori methods when one has no prior reasons to believe that one particular element will be different from the others. In this situation one would test the possibility that the largest of 765 communities was greater than the observed level. On the other hand, where one decides on the comparisons to be made before looking at the data, the investigator is entitled to use the much stronger method of a priori comparisons.

To gain an impression of the difference in the power between a priori and a posteriori comparisons it is sufficient to note that if the Sellafield result is considered as a priori comparison it has indeed a probability of only 1 in 10,000 while on a posteriori comparison has a probability of nearer 1 in 10, not even statistically significant at the standard .05 level. Which approach did the Black Committee use? Trying to read their minds, the claim that Millom Rural District's leukaemia rate was 'unusual but not unparalleled' appears to suggest that they are thinking in terms of extreme values, i.e. a posteriori comparisons, in effect. However there is another complication.

In the same way as one could envisage opportunistic browsing through the results of a number of communities to pick out the largest incidence, one could envisage a 'hostile observer' model where an opponent of (say) BNFL could compare the incidence of all possible diseases in all possible age groups between the area in question and the rest of the population and pick out the least favourable to Sellafield. (16) This procedure obviously bears a close relationship to the posteriori comparisons above but can be much more open-ended since there is not necessarily any pre-determined number of such comparisons, in contrast with the situation previously discussed where there was a pre-determined number of observations.

Which significance testing philosophy is appropriate?

At first sight it is not immediately clear whether to view the question as one of a priori or a posteriori comparisons. One might argue that the original YTV investigation aimed to investigate possible increases in radiation-related illnesses among the work force at Sellafield and that the enquiry was set up in response to fears arising from local observations of an increase in childhood leukaemia and that this represented 'peeking', so an a posteriori approach would be appropriate. However this does not hold water. We are not in a 'no knowledge' situation about the effects of nuclear radiation and it is perfectly legitimate, and indeed necessary, to use a priori approaches. If anyone feels that there is an element of hindsight in this, they should ask themselves one question. If you were asked to predict likely areas for an excess of radiation related illnesses, would you or would you not go for that around Sellafield, with a record of discharging quantities of radioactive materials roughly one hundred times as great as all the rest of Britain's nuclear plants put together (Black Report Figure 1.1). If it was going to be anywhere, it would be there. A priori rules ok?

The argument does not end at this point, however. The investigation considered leukaemia (or lymphoid malignancies) and all other cancers among children, as well as among adults for some investigations. It would be appropriate to take account of all these bites at the cherry by halving (or quartering as appropriate) the size of the test from .05 to .025 or .0125.

The net result of all these suggested changes in approaches is that the incidence of childhood leukaemia is statistically significantly greater around Sellafield than in other parts of Cumbria.

- 3) Internal Validity and third variables. No alternative 'third variable' hypothesis seems to have been put forward. The only alternative hypothesis which seems to have been put forward is that of a 'cluster' alias random fluctuation.
- 4) External Validity. On the one hand the external validity of the study is high in that the incidence in the Sellafield area is assessed on a comparable basis with the whole of the rest of England and Wales.

One could reasonably assume that the risk of excess childhood cancers in the environment of considerably cleaner nuclear plants would be vanishingly low, probably very low. In another way however, external validity is low in that the study provides little quantitative information on the link between radiation and increased risk of cancer.

- 5) Is the analysis performed at the correct level? There seems no reason to suggest that it is not. This question is more relevant to education research where results can be reported at individual level or on various aggregates - teacher, class, school, local authority.
- 6) Explanation of statistical techniques appropriate for the intended audience. In my opinion the Report does this very well.
- 7) Discrepancies in the care given to interpretation of different parts of the Report. The detailed statistical part of the Report tallies well with the general description of its findings.

8) Fabrication of the results. There is no evidence of this, and indeed, given that the Gardner & Winter data was collected for another purpose and tallies well with that collected by ITV, it is one of the strengths for the Report that one can be confident that the data is trustworthy.

Social Background

Research, social or medical, does not occur in a vacuum and it is of interest to speculate further on the basis of the five questions posed in Reading Between the Numbers to highlight the extent to which a piece of research is a social product.

1) Who commissioned and sponsored the research?

The study was commissioned and sponsored by the Government (DHSS) in response to public concern arising out of the ITV programme referred to earlier. The present Government has a stated policy of developing the nuclear industry, though there is no suggestion of any undue influence on the Committee: indeed the appointment of Sir Douglas Black as chairman reassured many people, myself included, that there would be no question of a cover-up. In fact the Sellafield question has ramifications considerably beyond judging one plant with a rather poor record of environmental discharges. One of the major practical difficulties with the peaceful uses of the nuclear energy is the disposal of the radioactive waste products. Sellafield was a reprocessing plant which not only earned valuable foreign exchange, but made the whole nuclear energy industry possible by dealing with the waste products. If it proved that it was impossible to carry this out safely, not only would Sellafield itself have to be shut down, with consequent loss of employment in an industrially depressed area, but the whole nuclear industry of the country would be in jeopardy, especially since Britain has failed in its attempts to relax the international ban on dumping nuclear waste at sea.

2) How was the research problem conceptualised and how well did the conceptualisation and the measures used fit together?

Perhaps mindful of the consequences of a decision that Sellafield represented a serious health risk, (or perhaps just out of habit) the Report opted for the standard Null Hypothesis model. This means that they had to be 95 per cent sure that Sellafield was a health risk before they reported it as such. In effect this means that the survival of the nuclear industry is much more important than that of those living near a nuclear plant. Different values of the significance level would have provided a different balance between those interests.

3) How was the research presented in its printed form and how was it received?

The Report itself was carefully presented accenting the qualified nature of the reassurance it offered and an emphasis on the provisional nature of the results and on the need for further research and monitoring of the situation. The nuclear lobby, perhaps predictably, acted as though the verdict, at best one of 'not proven' was one of 'not guilty'. The ITV participants reacted, as described earlier by sharp criticism of the Committee's results, in particular their apparent exclusion of some cases.

4) How was the research used and what interests were thus served? The results were used by the nuclear lobby, as outlined in (3) above to justify their contention that the nuclear power was indeed safe. British Nuclear Fuels Limited did however feel sufficiently exposed to mount an advertising campaign to publicise a hundred-fold reduction in emission of nuclear wastes, thus rather impaling themselves on Morton's fork: if it wasn't necessary, why were they doing it; if it was why hadn't they done it before?

5) Are there alternative uses of the research?

I have argued in this paper that the research paints an altogether more pessimistic view of the safety of nuclear reprocessing plants than Sir Douglas's qualified reassurance. It might well have been used as part of the grounds for a decision that nuclear power was too dangerous, and that the public resources employed therein be shifted to a search for other forms of energy. Or even to extending or subsidising coal mining.

Further Thoughts

- 1) Later cases. One reason that the nuclear industry is not justified in representing the results of the enquiry as a justification is that the Black Report recommended monitoring of the small area statistics around major nuclear installations. Some people ^{may} disagree with my contention that the stronger a priori methods of significance are more appropriate than a posteriori methods used by the Committee for the data up to 1982. However further cases will change the picture. Put somewhat colloquially we can say that the distinction is a question of 'pointing the finger' beforehand. While the Black Report argued that it was unable because of insufficient evidence, to say that there was an excess at the plant, it left the finger well and truly pointing at the area around Sellafield, (or whatever it's called this week). There can be no justification for not using a priori assessment for any subsequent incidence. In this connection, the further two cases of leukaemia which occurred in Sellafield are of particular importance. Two cases in two years (1983-4) are a very much more extreme incidence than 4 cases in fifteen years (1968-82), and the chance of their being due solely to some kind of non-systematic effect is vanishing small.
- 2) Finally it is worth emphasising the important role of the routine collection and publication and availability of statistical

information in putting issues into proportion. We have reached a situation in which rightly or wrongly, many people simply do not accept official reassurances on the safety of the nuclear industry. Apologists are viewed in some quarters at best naive and complacent and at worse as downright disingenuous. Some evidence for the former point of view, I fear, emerges from the radiobiological section of the Report itself.

In this situation the epidemiological information comes as an indispensable ray of light. It is conceivable that officially - published figures on radioactive discharges could contain quite serious underestimates. It is also quite likely that official safety levels for radiation dosage are much too high and will have to be revised drastically downwards. Similar doubts do not affect disinterestedly - compiled epidemiological information. The concerned outsider is thus able to form his or her own opinion on the basis of trustworthy evidence. At the risk of sounding like a commercial for the Central Statistical Office, this is something that is indispensable in a democracy.

Acknowledgements

Considerable thanks are due to Valerie Beral, Hazel Inskip, Peter Smith, Tony Ades, Ian Plewis, Liz Atkins, John Hodgson, Alison McFarlane and Andrew Davies for encouragement and prompt helpful comments which have improved the article significantly. Errors and opinions expressed are of course my own.

Dougall Hutchison

Notes

1. CRAFT, A.W., OPENSHAW, S. & BIRCH, J. (1984) 'Apparent clusters of childhood lymphoid malignancy in Northern England'. The Lancet 14/7/84 96-97
2. INDEPENDENT ADVISORY GROUP (1984) Investigation of the Possible Increased Incidence of Cancer in West Cumbria HMSO. (The Black Report).
3. New Scientist 1/11/84, 34-35
The Lancet 14/07/84, 95-96
New Scientist 22/11/84, 48-49
New Scientist 29/11/85, 45
The Lancet 4/08/84, 266-7
The Lancet 25/01/85, 172
The Lancet 19/01/85, 172
The Lancet 16/02/85, 403-407
New Statesman 18/01/85, 10-11
4. Black Report, 12
5. CUTLER, J.A. (1985) 'Checking the Figures' New Statesman 18/1/85 10-11

URQUART, J. & CUTLER, J.A. (1985) 'Incidence of Childhood Cancer in West Cumbria' The Lancet 19/1/85, 172.

6. Black Report, Table 2.18
CRAFT, A.W., OPENSHAW, S. & BIRCH, J. (1984) 'Apparent clusters of childhood lymphoid malignancy in Northern England'. The Lancet, 14/7/84
7. Craft & co-workers investigated all lymphoid malignancies while Gardner & Winter used leukaemia not including lymphomas. These are not exactly the same but there are sometime difficulties in distinguishing between them (Black Report Para. 2.49)

8. BIRCH, J.M. MARSDEN, H.B. SWINDELL R. (1980) Incidence of malignant disease in childhood: a 24-year review of the Manchester Children's Tumour Registry'. Br. Jr. Cancer 42 215-233
9. CRAFT, A.W. & OPENSHAW, S. (1985) Letter to the Editor Lancet I, 403
10. GARDNER, M.J., WINTER, J.D., TAYLOR, G.P. ACHESON, E.D. (1983) Atlas of Cancer Mortality in England and Wales. Wiley.
11. See note 5.
12. CRAFT, A.W. & OPENSHAW, S. (1985) Letter to the Editor, The Lancet, 16/2/85, 403-4
GARDNER, M.J. (1985) 'Letter to the Editor, The Lancet 16/2/85, 403.
13. See note 12.
14. See note 5.
15. See note 1.
16. Sir Richard Doll in his address to the Society for Social Medicine Annual Conference 1984 has identified a similar defect in some case-control hazardous substance studies.