Author XXX 3
Proof Reading Copy

J. Radiol. Prot. 1988 Vol. 8 No 1 00-00 Printed in the UK

Background radiation and childhood cancers†

E G Knox, A M Stewart, E A Gilmand and G W Kneale
Department of Social Medicine, University of Birmingham, Edgbaston, Birmingham B15 2TJ, UK

Received April 1987, in final form September 1987

Abstract. Outdoor terrestrial gamma radiation exposure levels (TGR), estimated for each of the 10 km squares of the Great Britain National Grid, were related to local cancer death rates in childhood. The examination was based upon the prior hypothesis that an association ought to be detectable. This was itself based upon an examination of geographical TGR variations and upon a recently reported recalculation of the dose–response relationship between the risk of childhood cancer and foetal exposure to medical x-rays. The analysis was pressed through several stages in which the effects of sociodemographic and medical confounding factors and their temporal changes were identified and separated. TGR was then shown to exert an independent statistically significant effect.

1. Introduction

The relationship between foetal x-ray exposure and the subsequent risk of childhood cancer in Great Britain was recently recalculated from the extended data base of a national case–control study[1]. The new estimates suggest a relative risk (RR) of 1.9 and a dose–response relationship of 2000 cancers per 10^4 Gy. For exposures in the first trimester the RR was substantially greater[2]. Over the past 30 years, when 11% of foetuses in the UK received diagnostic x-ray exposures, about 7% of all childhood cancers arose from this source[1].

Medical diagnostic procedures provide only part of the total radiation dose received by the foetal population. The remaining 'background' arises both from endogenous (^{40}K) and exogenous (cosmic, man-made and geological) sources. The geological sources include building materials and local rocks and both fractions are geographically variable. The extent of the geographical variation (see later), combined with the revised dose–response estimate for medical x-rays, suggested that an association between background radiation and childhood cancer should be detectable provided that the confounding effects of geographically varying social and medical determinants, including exposures to medical radiation, are eliminated. This approach therefore offered a prospect of confirming or amending the foetal dose–response relationships already derived, and that was the objective of this investigation.

Other geographical studies of background radiation exposure have been reported, but without resolving the question of a relationship. One Japanese investigation found an association between levels of background radiation and cancer deaths over 40 years of age [3] and another for cancer deaths in young children[4]. However, neither a follow-up of A-bomb survivors[5] nor an American nationwide study of background radiation produced any evidence of a cancer risk[6]. Investigations elsewhere have also been negative or equivocal [7–12], although in some cases because of small numbers or a narrow geographic variation, or because insufficiently sensitive methods of statistical analysis were used.

Availability of data: The authors are prepared to allow *bona fide* research workers on-site access to the data used in the analyses reported in this paper. A sub-set of the data will also be available shortly on magnetic media through direct application to the authors.

2. Materials and methods

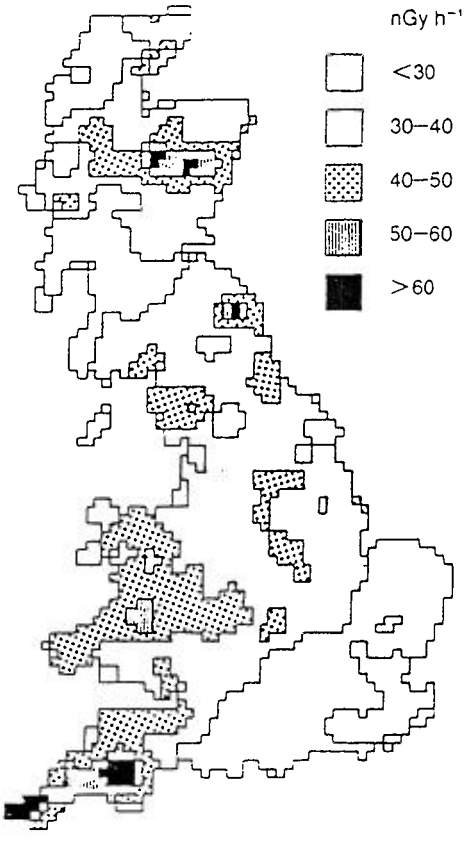
2.1 Data sources

The National Radiological Protection Board (NRPB) recently completed a nationwide

0260-2814/88/010000 + S02.50 © 1988 IOP Publishing Ltd.

(England, Scotland and Wales) survey of terrestrial gamma radiation (TGR), taking one or more measurements in the open air in 95% of the 2400 10 km grid squares of Great Britain. Where more than one measurement was made in a square, the mean was calculated, and for squares where no measurement was made (mainly because of their remoteness) the value was estimated from data for surrounding squares. Because individual measurements have a high intrinsic variability, a two-dimensional 'moving mean' was then calculated, the measurement for each square being modified using values from adjacent squares[13]. The NRPB kindly made this material available to us in advance of publication, together with a map illustrating the form of the (smoothed) distribution of TGR across England, Wales and Scotland (figure 1).

Figure 1: Terrestrial gamma-ray dose rates out-of-doors in Great Britain
(Note: Cosmic and intrinsic backgrounds removed (Data smoothed twice))



From these data we calculated that the average absorbed dose rate in air was approximately 34 nGy h^{-1}, varying from about 15 to 60 nGy h^{-1}. During a foetal life of 270 d this accumulates to 0.22 mGy. This is only 40% of the average received dose of 0.55 mGy from medical radiation (5 mGy per examination in 11% of the population) during the period 1944–1979[1]. The cumulative 'outdoor' TGR exposures (between 0.1 and 0.4 mGy) show less individual variation than exposure or non-exposure to medical x-ray examination, but TGR covers the early and more radiosensitive stages of gestation; thus the range of carcinogenic effects from the two sources might be comparable.

In addition, TGR measurements represent a wider variation of exposure than the raw TGR estimates at first suggest. The pregnant woman spends more time indoors than outside. Indoor radiation exposures are greater than outdoor exposures by a factor of 1.8 and the building materials from which they emanate (in part) are often, although not always, from local sources[14]. Systematic national data on indoor exposures are not yet available, but to the extent that indoor and outdoor exposures will eventually be shown to be correlated, the outdoor TGR variations must reflect a larger absolute dose variation than the raw figures themselves suggest.

The Oxford Survey of Childhood Cancers (OSCC), from which the medical radiation measurements were derived, is a case–control study of all children dying from cancer in the UK between 1953 and 1979 and born between 1944 and 1979. For children born between 1953 and 1964, there was complete ascertainment of all cancer deaths between birth and 16 years of age, while for cohorts born outside this period the age limits of the follow-up are truncated in early or late childhood. The analysis was based upon 22 351 children dying from cancer.

All death addresses (or birth addresses where these were known to be different) were coded to one of 1797 local authority districts (i.e. metropolitan, county and municipal boroughs, urban areas and rural districts which existed in 1965), for which annual numbers of live births were also available.

2.2 Method of analysis

In order to align the OSCC data with the TGR

values, each local authority district was assigned to a National Grid 10 km square. The appropriate squares for boroughs and urban districts were assigned directly from the index of the *Ordnance Survey Motoring Atlas*. When rural districts had the same names as boroughs or urban districts, the two were merged and linked jointly with the location of the urban district. This resulted in a condensed map of births, cancers and TGRs in which the original number of ~~1791~~ local authority districts was reduced to 911 TGR-related demographic districts (DD). The 911 DDs were located within larger grid squares (100 km) which are subsequently referred to as demographic regions (DR). Cancer death rates for each DD were calculated by relating numbers of cancer deaths in component local authority districts to numbers of live births in each annual birth cohort over a period of 36 years from 1944 to 1979, and to numbers of years of follow-up attained for each of these birth cohorts. The 36 birth years and the 911 DDs supply 32 796 individual DD cohorts with follow-up periods ranging from 1 to 16 years. The total surveyed experience amounted to 347 564 836 child years.

Each cancer death had been allocated a healthy control from the birth file of the district address at the time of death. The recalculation of the effects of medical radiation already reported was based upon case–control contrasts among the 14 759 matched and interviewed pairs for which a full set of social, medical and demographic data was available. Because the controls were matched in terms of location at death as well as in terms of sex and date of birth, the paired control data could not themselves indicate any differential geographical distribution of the deaths. This was to be based primarily upon a regression analysis between the cancer death rate and the TGR measurement in each DD.

However, the controls supplied additional data on geographical and temporal distributions of social, medical and demographic factors on which childhood cancer death rates are known partly to depend, as well as providing estimates of the extent to which the risk was altered by their presence, and this information provided a basis for standardising for these effects and for their geographical distributions. Because of the temporal component, the analyses involving control data were based upon individual single-year DD cohorts rather than aggregated DD death rates over the full 36 year period. These more detailed analyses were based upon the place of birth of cancer cases rather than place of death; the two differed in 16.4% of the cases interviewed.

The analysis of cancer risk in relation to TGR values was conducted in three separate stages:

(i) A simple regression analysis of death rates upon TGR was carried out for the 911 DDs over the full 36 year period with no attempt to standardise for confounding factors.

(ii) A more complex regression analysis was carried out jointly at a coarse and at a fine level, again over the full 36 year period. The coarse level of analysis used DR-aggregated death rates and mean TGRs, while the 'fine' (DD) analysis was repeated *within* each of these larger squares. The effects of local TGR variations, within DRs, were thus separated from general geographic trends such as might be determined by gross social variations.

(iii) Finally, a regression analysis of death rates in relation to TGR (and a number of other factors) was carried out using conditional logistic regression (Miettinen–Breslow) techniques. The details of the method are described in appendix 1. Systematic sociogeographic trends were eliminated through incorporating Northing and Easting parameters, and data from the controls were used in order to eliminate the residual effects of social and health-care variables (including the frequency of medical radiation) not expressed in latitude or longitude. This detailed analysis was based upon separate DD cohorts rather than the year-aggregated DDs used in the first two stages. Not all DD

Table 1. Cancer death rates for eight levels of TGR.

TGR dose rate (nGy h^{-1})	Person years of follow-up	Cancer deaths	
		No.	Rate per 10^5 person years
12–19	5 103 116	362	7.09
20–24	45 687 955	2973	6.51
25–29	87 605 332	5664	6.47
30–34	69 027 813	4334	6.28
35–39	73 368 981	4826	6.58
40–44	56 315 974	3548	6.30
45–49	4 399 802	255	5.80
50–82	6 055 863	389	6.42
12–82	347 564 836	22351	6.43

cohorts (and indeed not all DDs) were 'occupied' by cases and controls; only 9943 of the 32 796 DD cohorts contained one or more cancer deaths.

3. Results

3.1 Stage 1

The joint distribution of TGRs and death rates in the 911 DDs are given in table 1. There was no evidence of a direct radiation effect. Indeed, the lowest TGR dose band (under 20 nGy h^{-1}) had the highest cancer rate, and the second highest dose band had the lowest. Aggregation of the National Grid into three zones running from east to west is demonstrated in table 2. This also shows a tendency for the death rate to run *against* the gradient of background radiation.

3.2 Stage 2

Table 3 lists 36 DRs in descending order of person years of surveillance, and it records the numbers of 'occupied' DDs within each of them. (Four DRs with only a single occupied DD are omitted.) Person years of experience, number of cancers and ranges for TGR are given. The final two columns give the linear coefficients of dependency *(b)* of DD cancer rates upon TGRs *within* each DR and *t* values. Three of the twenty positive coefficients were statistically significant: five if we accept a single-tailed criterion. Three of these five were in the three most populous regions. None of the negative coefficients was significant. The aggregated coefficient for DD variation within all DRs together was positive. It was statistically significant on a single-tailed criterion ($t = 1.83$).

These results do not permit a clear-cut conclusion, but they differ from those of the stage 1 analysis. This suggests that there might indeed be a general covariation between cancer and TGR, but one that is dominated by contrary geographical gradients of the confounding social determinants.

3.3 Stage 3

The third analysis was designed to test this hypothesis. Unlike the earlier analyses, it was based upon individual DD cohorts defined in a year-of-birth as well as in a geographical dimension, and with cancer cases located according to their place of birth rather than their place of death. First, however, it was necessary to explore the question of space–time heterogeneity. This is partly because it is an important contemporary question in its own right, but also because the question of statistical significance hinges upon the premise that individual occurrences are independent of each other. The question of clustering was first investigated through deriving distributions of numbers of cancer deaths, and expected distributions, among different DD cohorts classified according to their population-determined mean expectations. A condensed representation of these distributions is given in table 4. The overall expected distribution is the sum of the Poisson expansions for each row. It can be seen that except for the last column the expected and observed distributions are similar.

A χ^2 test of heterogeneity revealed a statistically significant non-uniformity between the DD cohorts but it was not sufficient to reveal itself in the tabulated distribution. When χ^2 was partitioned into its spatial, temporal and space–time components, the variations were shown to be entirely spatial and temporal, with no evidence of an interaction between them. At this level of temporal and spatial resolution, therefore, there was no evidence of space–time

Table 2. Cancer death rates and TGR dose rates for geographical zones.

Zones	Person years of follow-up	Cancer deaths		TGR dose rate (nGy h^{-1})
		No	Rate per 10^5 person years	
West	69 731 256	4281	6.14	37.1
Centre	145 543 857	9371	6.44	34.0
East	132 289 723	8699	6.58	31.3
All regions	347 564 836	22351	6.43	33.9

Table 3. Cancer rates and TGR exposures within 36 demographic regions.

Demographic regions[†] Code	No of DDs	Person years follow-up per DD Rank	Average 10^3	Cancer deaths 1953–79	TGR dose rates (nGy h^{-1}) Lowest	Highest	Regression coefficients[‡] β	t
TQ 51	61	1	1157	4685 4624	21	35	+ 1.30	4.00§
SJ 33	53	2	686	2207	29	42	+ 0.87	1.94§
NZ 45	29	3	646	1109	30	45	+ 1.19	1.91§
NS 26	39	4	569	1281	23	35	+ 0.07	0.07
SP 42	53	5	469	1649	23	48	− 0.33	0.97
SK 43	54	6	457	1587	30	45	− 0.63	1.48
SE 44	49	7	418	1318	28	43	− 0.71	1.30
SD 34	42	8	408	1075	25	44	+ 0.95	1.08
SU 41	47	9	370	1241	13	37	− 0.07	0.13
TA 54	15	10	351	318	28	39	+ 2.33	2.07§
ST 31	50	11	293	928	21	46	− 0.46	0.92
SO 32	48	12	267	838	35	49	+ 0.96	1.15
TL 52	47	13	258	896	21	34	+ 0.50	0.51
SX 20	19	14	238	278	37	48	− 0.85	1.20
SS 21	20	15	237	258 268	32	48	+ 0.70	0.88
SZ 40	12	16	230	159	12	48	+ 0.19	0.18
NT 36	31	17	202	416	27	30	+ 0.29	0.35
TR 61	13	18	195	184	25	35	+ 1.38	0.87
TG 63	11	19	180	142	20	28	− 4.05	1.03
NO 37	24	20	156	258	18	46	+ 1.02	0.58
NJ 38	19	21	153	190	30	51	− 0.26	0.51
TM 62	26	22	145	255	22	29	− 1.17	1.04
SM 12	4	23	143	46	43	45	+ 2.99	0.13
NH 28	8	24	131	70	29	46	− 0.66	0.29
NY 35	17	25	130	138	30	47	− 1.06	1.05
NR 16	2	26	117	14	24	29	−13.39	0.76
TF 53	26	27	116	213	21	40	+ 0.15	0.18
SY 30	9	28	115	90	19	41	+ 1.58	1.33
SW 10	10	29	113	92	47	82	+ 1.03	1.47
NX 25	11	30	82	82	22	45	+ 0.09	0.06
SN 22	18	31	84	111	30	49	− 0.04	0.03
NN 27	6	32	79	33	32	39	+27.02	2.89§
NC 29	3	33	68	10	29	37	+ 3.81	0.75
SH 23	24	34	56	89	24	50	− 1.13	0.61
NU 46	5	35	49	24	41	48	− 3.66	0.39
NM 17	2	36	24	8	18	34	− 4.27	0.33
Total	907	Total	378	22 301	12	82	+ 0.23	1.83§

[†] Excluding four DRs where there was no variation of TGR at district level.

[‡] Cases per 10^6 per (nGy h^{-1}).

§ $P < 0.05$ (one-tailed).

interaction: that is, no evidence of geographical 'concentrations' and 'gaps' occurring in different places at different times.

Regression calculations of death rate upon TGR for different DD cohorts were next performed within the framework of the Miettinen and Breslow (conditional logistic regression) procedure for analysing matched pairs. This was done by incorporating the number of cancer deaths and the number of person years of surveillance within the appropriate DD cohort, as additional variables of each case and control. This analysis was performed upon the same subset of the population (14 759 case–control pairs) as that from which the medical radiation dose–response relationships had been

Table 4. Distribution of cohorts according to observed and expected numbers of cancer deaths.

Expected no of deaths	Observed no of deaths											Total cohorts
	0	1	2	3	4	5	6	7	8	9	10+	
0.0–0.5	20 379	3254	465	74	14	1	0	0	0	0	0	24 187
0.6–1.5	2 503	2021	889	315	93	19	9	1	0	0	0	5 850
1.6–2.5	209	377	341	216	128	71	17	6	4	0	1	1 370
2.6–3.5	37	90	134	105	66	50	24	16	1	5	4	532
3.6–4.5	5	29	51	64	65	46	37	18	9	6	6	336
4.6–5.5	2	5	28	19	25	28	22	22	15	13	8	187
5.6–6.5	2	2	7	9	19	27	32	19	19	8	19	163
6.6–7.5	1	0	4	4	9	20	13	17	13	10	26	117
7.6–8.5	0	1	1	1	3	7	9	5	12	11	23	73
8.6–9.5	0	0	1	2	0	3	7	3	4	9	13	42
9.6–9.9	0	0	0	1	1	1	1	1	1	1	6	13
10+	0	0	1	3	0	2	9	8	3	10	171	207
Total												
Cohorts[a]	23 138	5779	1922	813	423	275	180	116	81	73	277	33 077
Expected[b]	22 961.1	5963.8	1939.0	827.6	434.7	267.7	184.8	135.8	104.3	79.4	177.8	33 077.0

[a] All cohorts with births >0, including 281 cohorts from nine 10 km squares for which no TGR data were available.

[b] The expected distribution was calculated as the sum of the Poisson expansion for each individual row of the original table.

calculated. This incurred a risk that cancers for which full data were available might not accurately represent the totality of childhood cancer deaths; there was also some loss of information in that the DD cohorts which truly yielded no cancers at all could not contribute to the analysis. This also demanded modification of significance-testing procedures for regression coefficients, because the 'expected' distribution of cases had to be treated as a zero-suppressed, rather than a full Poisson distribution. However, the advantages of studying the effects of TGR dependency and of the sociodemographic and medical variables within a single procedure are plain enough. A technical description of the method is given in appendix 1.

Table 5 gives the results of a conditional logistic regression analysis of cancer death rates upon TGR among the 9943 DD cohorts which qualified for entry. The first panel extracts variation attributable to sociodemographic factors, of which maternal age showed the only statistically sigificant effect. The second panel shows the significant effect of prenatal medical x-rays. There is a significant interaction between x-rays and year of birth, a reflection of the declining carcinogenicity of medical x-ray examination over this period of time[1]. The type of cancer (factor 7) made no difference: the determinants in this table affected the risk of the leaukaemias and of the solid cancers equally. The final panel demonstrates the independent statistically significant effects of geographical location and of TGR. Once more, the effect was independent (factor 16) of the type of neoplasm.

The unit of change (i.e. the interval between 'levels') for pre-natal x-ray was about 5.0 mGy while for TGR it was defined in the terms of the NRPB data as 1.0 nGy h^{-1}. The ratio between the respective regression coefficients (β), 0.7413/0.0034, shows that a single radiological examination has the same effect as about 218 TGR units, or, accumulated over the full gestation period of 270 d, about 1.4 mGy (i.e. 218 units × 270 d × 24 h × 10^{-9} Gy). TGR therefore appears to be about 3.6 times (5.0/1.4) as effective, dose for dose, as medical x-ray exposure.

This difference might reflect the concentration of pre-natal x-rays in the third trimester, compared with the wider distribution of TGR across all trimesters, including the particularly radiosensitive first trimester[2]. It could also reflect the likelihood that geographical variation of TGR reflects the correlated variation of a larger fraction of background radiation than that represented by the out-of-doors TGR estimate alone. However, the estimated ratio (3.6) has rather wide confidence limits, and there are additional grounds for treating such calculations with caution.

Table 5. Results of the regression analysis of cancer death rates upon TGR.

Factor (units)	βx ‡	t‡
1. Maternal age (years)	−0.0079	+4.94 ***
2. Sex	−0.0138	−0.81
3. Sibship position	−0.0041	−0.71
4. Social class (I–IV)	−0.0198	−1.92
5. Year of birth	0.0010	+0.77
6. Pre-natal x-ray	0.7413	+3.52 ***
7. Cancer Diagnosis (RES or solid)	0.0507	+0.66
8. Cancer onset age (months)	0.0003	+1.37
9. (Cancer onset age)^2	-3.40×10^{-6}	−0.69
10. (x-ray) × (onset age)	−0.0004	−0.84
11. (x-ray) × (onset age)^2	-5.35×10^{-6}	−0.62
12. (x-ray) × (year of birth)	−0.0085	−2.55 *
13. Place of birth: Easting (10 km)	0.0047	+4.05 ***
14. Place of birth: Northing (10 km)	−0.0012	−1.86
15. TGR dose rate (nGy h^{-1})	0.0034	+1.96 *
16. (TGR dose rate) × (diagnosis)	−0.0012	−0.51

‡ βx: change in ϑn (relative risk) per unit change of each factor.

‡ Levels of significance: * $0.01 < p < 0.05$; ** $0.001 < p < 0.01$; *** $p < 0.001$.

The London conurbation presents us with a difficult problem in that it has a low TGR and, in common with several other large urban areas, a low cancer death rate. London is so large that it might alone be responsible for the statistical association between TGR and cancer death rate, and for reasons other than a direct effect of radiation itself. Fortuitous urban/rural associations with local TGR measurements, elsewhere in the country, could also contribute towards a spurious result. We therefore incorporated the binary urban/rural classification of the original civil district as a confounding variable. Table 6 shows the independent effects of urban status and of TGR.

4. Discussion

This investigation was designed to test the prior hypothesis, based on previous evidence, that an association between childhood cancer death rates and background radiation should be detectable within the UK. The hypothesis followed from a recent recalculation of dose–response relationships for foetal x-ray exposure with respect to childhood cancer, and examination of the results of a national terrestrial gamma radiation (TGR) survey carried out by NRPB across the whole of England, Wales and Scotland. We deduced that if the new estimates of the dose–response relationship for x-rays were correct, then a geographical relationship between TGR and cancer incidence should be demonstrable. This would probably require simultaneous allowance for a number of confounding factors also known to influence the risk, but this could be achieved if the background radiation data were combined with the sociodemographic and medical data already to hand.

In the event, simple regression analysis of childhood cancer death rates upon TGR across the 10 km squares of the National Grid failed to detect such a relationship, although a second level of analysis, using both a coarse (100 km square) and a fine (10 km square) disaggregation of death rates and TGRs, looked more promising. The second analysis suggested that nationwide gradients of social and demographic factors might be masking a true TGR effect, so that it was evident only within socially homogeneous zones. A conditional logistic regression analysis, taking simultaneous account of sociodemographic and medical-care variations in different parts of the country and

Table 6. Results of the regression analysis including urban/rural status as an independent variable.

Factor (units)	$\beta x'$	$t^{\ddagger}$
1. Maternal age (years)	−0.0078	+4.89 ***
2. Sex	−0.0141	−0.83
3. Sibship position	−0.0040	−0.69
4. Social class (I–IV)	−0.0193	−1.87
5. Year of birth	0.0009	+0.74
6. Pre-natal x-ray	0.7393	+3.51 ***
7. Cancer Diagnosis (RES or solid)	0.0003	+1.35
8. Cancer onset age (months)	-3.24×10^{-6}	−0.92
9. (Cancer onset age)^2	−0.0004	−0.85
10. (x-ray) × (onset age)	-5.20×10^{-6}	−0.60
11. (x-ray) × (onset age)^2	−0.0085	−2.54 *
12. (x-ray) × (year of birth)	0.0126	−0.74
13. Place of birth: Easting (10 km)	0.0047	+4.05 ***
14. Place of birth: Northing (10 km)	−0.0011	−1.76
15. TGR dose rate (nGy h^{-1})	0.0432	−2.29 *
16. Urban$^{(1)}$ or rural$^{(0)}$	−0.0026	+2.08 *

β = $\beta x'$: change in ϑn (relative risk) per unit change of each factor.

‡ Levels of significance: * $0.01 < p < 0.05$; ** $0.001 < p < 0.01$; *** $p < 0.001$.

in different cohorts then revealed a significant TGR relationship.

The risk per unit dose from TGR was greater than from pre-natal x-rays, and this accorded with prior expectation. TGR exposure occurs in all trimesters, including the very sensitive first trimester, while pre-natal x-rays are largely concentrated in the third trimester. There is a second contributory explanation: TGR is only one component of the geographically variable fraction of total background radiation, and since the different components are likely to be positively correlated, TGR values must represent greater absolute exposure variations than the measurements themselves suggest.

Such quantitative estimates must be treated with caution. We can probably rely upon the qualitative demonstration of a relationship between TGR and cancer risk, but the effects of fortuitous non-independent geographical clusters of the social determinants and the geological formations may have perturbed the quantitative outcomes. However, to the extent that it has been possible to press the analysis, these outcomes do confirm the prior hypothesis established at the outset.

We have calculated that pre-natal x-ray exposures accounted for about 7% of all childhood cancers over the past 30 years[1]. The accumulated foetal exposure to background radiation from all sources over the full gestation period is several times greater than the average dose (exposed plus non-exposed) from pre-natal x-rays. The background dose is also received at a more sensitive foetal age. These facts, together with the demonstration of a TGR effect, suggest that radiation might be an element of the causal chain in the majority of childhood cancers. Proportional increases in overall foetal radiation exposures, from whatever cause, would then be expected to result in a near-proportional increase in the subsequent cancer rate.

The Chernobyl incident might provide a means of testing this prediction and consolidating our knowledge of the dose–response relationship. The appropriate technique would be to accumulate dates of birth of childhood cancers as they arise, and relate the size of the birth-date step (around November 1986) to different exposures recorded in different localities.

Acknowledgments

This study was supported by the Radiation Protection Programme, Contract No B10-F-541-83-UK, of the Directorate-General for Research, Science and Education, Commission of the European Communities. The programme of which this study is a part was also supported by the MRC, Contract No G812-605-7, and by the US Department of Health and Human Services, Contract No 233-78-6017.

Appendix 1

A.1. Modified Miettinen–Breslow analysis

The application of standard conditional logistic regression techniques to the matched-pair data used in this analysis were reported in a previous paper[1]. In the present analysis it was necessary to incorporate two new variables, namely TGR and death rate, which (except for those children who moved house between birth and death) were common to both cases and controls. The analysis therefore took the form of a logistic regression analysis of the dependency of death rate upon both TGR and a number of other factors, most of which also supplied case–control comparisons. The method was adapted to standardise for two main classes of confounding variable namely *(a)* the geographically distributed values for the controls alone, and *(b)* the case–control contrasts and their interactions with cancer risk. The analysis proceeded as follows.

(1) Let x_i be a vector of observed values for case i, and let y_i be the vector of corresponding values for control i in a situation where x_i equals y_i for all matching or case-only factors.

(2) Let β be a vector of regression coefficients and α a constant such that the estimated probability of case i being an actual case $(p_i) = \exp(\alpha + \beta\, x_i)$ and the corresponding probability for control i is $(q_i) = \exp(\alpha + \beta\, y_i)$.

In these circumstances α will equal the baseline logarithm of case incidence. Then, according to Miettinen (and with the stated values of x_i *and* y_i), the conditional probability of any case/control pair having the values actually observed is $p_i/(p_i + q_i)$, and, according to Breslow, the best estimates of β are given by maximising the log-likelihood

$$L_1 = \Sigma_i \vartheta n(1 + \exp[\beta\,(y_i - x_i)])$$

i.e. the standard conditional logistic regression formula.

A.2. Application to OSCC data

(1) Let there be N_r cases for DD cohort r, and let the total person years of follow-up be T_r. Then the expected incidence is $Q_r = (\Sigma_{ir}\, q_i)/N_r$ where Σ_{ir} means summation over N_r case/control pairs for DD cohort r.

In practice it is values of matched controls rather than the values of cancer cases which feature in this formula since the controls should be a random sample of the regional birth cohort and therefore more typical of DD cohorts than cases.

The expected value of N_r *(*M_r*) is* $T_r\, Q_r$, and the variance of N_r about this value is proportional to M_r. Therefore, best estimates of α and β are obtained by minimising any discrepancy between N_r and M_r and this discrepancy $Z_1 = \Sigma_r (N_r - M_r)^2 / M_r$.

This expression is similar to a χ^2 formula, whereas the expression for L_1 (above) is a log-likelihood. Therefore, an an alternative measure of discrepancy we have $Z_2 = \Sigma_r\, N_r\, \vartheta n(M_r) - M_r$ since this is the variable part of the log-likelihood ratio statistic corresponding to Z_1 as a χ^2 statistic.

As an expression to be evaluated Z_2 has two disadvantages: in the first place Z_2 depends upon control values. Therefore, it cannot be evaluated for DD cohorts with no cancer cases (and therefore no matched controls). Secondly, although L_1 can be expressed as a sum over all case-control pairs this is not possible with Z_2.

The first problem was solved by treating N_r (and its expectation M_r) not as an ordinary Poisson variable (as in the conventional χ^2 statistic derivation by Fisher) but as a zero-suppressed Poisson variable. This meant adding an additional term to Z_2 and thus converting it to

$$Z_3 = \Sigma_r (N_r\, \vartheta n(M_r) - M_r - \vartheta n(1 - \exp(-M_r))).$$

The second problem was solved by realising that when βy_i *is small compared with* α, any function $f(M_r)$ *of* M_r will be approximately equal to

$\Sigma_{ir} (f(N_r m_i))/N_r$ where $m = T_r\, q_i / N_r$ is the contribution of case i to M_r. Thus, we finally have an approximate log-likelihood

$$L_2 = -\Sigma_i \left(m_i \ln(m_i) + \ln(1-\exp(-N_i m_i))/N_i\right)$$

which may be added to L_1 and the whole maximised for variation in α and β.

Finally, the statistical analysis recognised that there might be missing data, thus making it necessary to give median values for certain items such as sibship (second), maternal age (25–29 years) and social class (grade III).

A.3. Unsolved technical questions

There are several unsolved questions requiring resolution before the above methods can be described as definitive.

(1) In the first approximation (Z_1) there are two sources of variance. The first is the Poisson variation of N_i about M_i; this is proportional to M_i. Also there is variation of M_i about its true value. This variation comes about because there are only N_i controls to estimate the true mean value of Q_i. Fortunately, this extra variation is proportional to N_i and may therefore be taken as proportional to M_i when N_i is allowed to vary. Thus the total deviance of Z_1 or Z_2 will be multiplied by a constant factor. This constant can be estimated as follows:

It can be shown that L_2 achieves its maximum value, with independent variation of the m_i, when m_i is equal to e_i, with $e_i = 1-\exp(-N_i e_i)$. This maximum value is given by

$$C = -\Sigma_i \left[e_i - (1-1/N_i)\ln(e_i)\right].$$

Thus the deviance can be estimated by $V = -2(L_2 - C)$. A possible query is the number of degrees of freedom of this estimate. However, on a typical analysis with 21 factors, 22 351 cancers and 9943 cohorts with at least one cancer death, the value of V was 9324.99. Therefore, the deviance is probably less than any likely number of degrees of freedom, and equal weighting of L_1 and L_2 would seem to be appropriate.

(2) The second query concerns the optimality of any linear combination of L_1 and L_2 in estimating α and β. Strictly speaking, this could only be justified if L_1 and L_2 were independent. However, considerations of orthogonality suggest that, since L_1 (the Miettinen–Breslow log-likelihood) is strictly a function of the difference between case and control values, the corresponding function, L_2, in the regression analysis ought perhaps to be strictly a function of the sum of case and control values, rather than just control values only; however, the correct formula for this function is not obvious.

References

[1] Knox E G, Stewart A M, Kneale G W and Gilman E A 1987 Prenatal irradiation and childhood cancer *J. of the Soc. for Radiol. Prot.* in press

[2] Gilman E A, Kneale G W, Knox E G and Stewart A M 1987 Pregnancy x-rays and childhood cancer: effects of exposure age and radiation dose *J. of the Soc. for Radiol. Prot.* in press

[3] Ujeno Y 1978 Carcinogenic hazard from natural background radiation in Japan *J. Radiat. Res (Tokyo)* **19** 205-12

[4] Sakka M 1979 Background radiation and childhood cancer mortality *Nippon Acta Radiol.* **39** 536-9

[5] Kato H and Schull W J 1980 Studies of the mortality of A-bomb survivors *Report 9. Mortality, 1950-1978: Part 1* RERF Life Span Study, RERF-TR pp 12-80

[6] Frigerio N A, Echerman K E and Stowe R S 1973 The Argonne Radiologic Impact Program Part 1. Carcinogenic hazard from low-level, low-rate radiation. ANL/ES-26, Part 1

[7] Court-Brown W M, Spiers F W, Doll R *et al* 1960 Geographical variation in leukaemia mortality in relation to background radiation and other factors *B. Med. J.* **1** 1753-9

[8] Mason T J and Miller R W 1974 Cosmic radiation at high altitudes and US cancer mortality, 1950-1969 *Radiat. Res.* **60** 302-6

[9] Pincet J and Masse L 1975 Natural radiation and cancer mortality in several areas of North Brittany *Int. J. Epidemiol.* **4** 311-6

[10] Jacobson A P, Plato P A and Frigerio N A 1976 The role of natural radiations in human leukemogenesis. *Am. J. Public Health* **66** 31-7

[11] Hickey R J, Bowers E J, Spence D E *et al* 1981 Low level ionizing radiation and human mortality: multi-regional epidemiological studies *Health Phys.* **40** 625-41

[12] Allwright S P A, Colgan P A, McAulay I R and Mullins E 1983 Natural background radiation and cancer mortality in the Republic of Ireland *Int. J. Epidemiol.* **12** 414-8

[13] Green B M R, Lomas P R, Bradley E J and Wrixon A D Outdoor gamma ray dose rates in Great Britain National Radiological Protection Board, Harwell, Report No. NRPB R-191 in preparation

[14] Hughes J S and Roberts G C 1984 The radiation exposure of the UK population — 1984 review National Radiological Protection Board, Harwell, Report No. NRPB R-173 (London: HMSO)