COMARE PAPERS **COMARE 12-20**

FOR MEMBERS' USE ONLY

Meeting: **Agenda No: 4.i**

COMARE 103 **10th October 2012**

For information and discussion

Paper enclosed:

Preliminary Report on Cancer Rates around Dalgety Bay.

Author(s):

Preliminary Report on Cancer Rates around Dalgety Bay

1. Introduction

Following the reports of radioactive contamination around the headland at Dalgety Bay, the Information Services Division of the NHS National Services Scotland were asked to investigate the incidence of cancer in the area. In COMARE 12-02 (18 March 2012) rates in the vicinity of the headland for the years 2000-9 were reported in a table showing 27 tumour groups, divided into groups according to prior knowledge about their possible radiation aetiology. The table reports 99% confidence intervals, using a higher than usual confidence coefficient to reflect the fact that several tests were carried out. The only tumour amongst those with some established association with radiation was liver cancer, with 10 cases observed out of 4.0 expected. In discussion with

I offered to do a spatial analysis of the data on liver cancer; the results of this are presented in §3 after an assessment in §2 of the extent to which the table as a whole shows evidence of raised risk levels.

2. Overall Incidence Rates

Table 1 below shows the original data for all 27 tumours, in layers corresponding to prior knowledge about a possible aetiology, as described in COMARE 12-02. An extra column has been added to show the "deviance", a measure of discrepancy between observed and expected numbers that gives a better picture than the P-value (which may be misleading with very small numbers), or the SIR (Obs/Exp, which can be disproportionately large when expectations are small). With larger expectations it behaves rather like the familiar Pearson's chi-square, but is much less prone to extreme values with smaller frequencies.

Although the number of liver cancers is not the most conspicuous result in the whole table by any of the criteria in the right-hand three columns, it is so if we restrict attention to those tumours with "robust" evidence of a radiation aetiology – the first two divisions of the table. But given that is it one of 11 comparisons made in these divisions, it is of interest to determine whether its deviation from expectation is larger than we might expect amongst this number of them.

To investigate this, artificial data were simulated for the 11 tumours in the top half of the table under the null hypothesis that the observed numbers are independently Poisson-distributed with the expectations calculated. Out of 1000 such simulations, 54 returned a maximum deviance greater than that observed for the real data (6.33). This would suggest a result almost significant at the 5% level after making adjustment for the multiple testing. This test is one of a number that could be performed and, if we compute instead the simulated P-values, we find only 17 of the simulated values less than the value for the observed data (0.0081). This suggests a rather more significant result of around 2%. In the absence of a generally agreed best method to use, we would perhaps conclude that there is evidence of an excess in the liver cancers at something like the 5% level after adjusting for multiple testing.

	Tumour	Obs	Exp	SIR	P-value	Deviance
1	Myeloid	6	4.9	1.23	0.37	0.23
2	Thyroid	3	2.9	1.03	0.55	0.003
3	Breast	61	73.8	0.83	0.94	2.36
4	Lung	41	43.9	0.93	0.69	020
5	Stomach	6	10	0.6	0.93	1.87
6	Colon	34	38.9	0.87	0.80	0.64
7	Oesophageal	6	10.4	0.58	0.95	2.20
8	Bladder	9	10.4	0.86	0.71	0.20
9	Ovary	9	10.7	0.84	0.74	0.29
10	Brain/CNS	6	7.7	0.78	0.78	0.41
11	Liver	10	4	2.47	0.0081	6.33
12	Kidney	8	10.2	0.78	0.80	0.51
13	Salivary	2	0.5	3.65	0.090	2.55
14	NHL	27	15.8	1.71	0.0064	6.53
15	Myeloma	5	6.1	0.82	0.73	0.21
16	Skin	20	22.7	0.88	0.74	0.33
17	Rectum	18	19.9	0.9	0.70	0.19
18	Uterus	12	9.7	1.23	0.27	0.51
19	Bone	0	0.4	0	1.00	0.80
20	Connective Tissue	2	2.9	0.69	0.79	0.31
21	CLL	7	4.6	1.51	0.18	1.08
22	Pancreas	12	9.3	1.28	0.23	0.71
23	Hodgkin Lymphoma	7	2.6	2.68	0.020	5.07
24	Prostate	61	52.4	1.17	0.13	1.34
25	Testis	3	4.2	0.71	0.79	0.38
26	Cervix	3	3.6	0.83	0.70	0.11
27	Selected Childhood	0	0.4	0	1.00	0.80

Table 1. Observed and expected numbers of cases of cancer in the Dalgety Bay area, 2000-09, with SIR, P-value and deviance. The table sections are in decreasing order of prior evidence for a radiation aetiology.

If we look at the table as a whole, we find that neither the smallest P-value nor the largest deviance is especially extreme, given that we have tested 27 tumours. However, the position is complicated by the fact that more than one tumour shows a discrepancy and this would suggest a different kind of test. But here we are in danger of using our knowledge of the results to select a test. Formal statistical inference is running into difficulty, but we conclude that both the liver cancers and the NHL might be worth a second look. Possibilities for further investigation include looking at earlier data and examining the spatial distribution of the cases of a particular tumour and this we address for liver cancers in §3.

3. Spatial distribution of liver cancers

kindly supplied spatial data in the form of areal counts in the 13 data zones surrounding the Dalgety Bay site. The distances to the headland were calculated and used for a test of proximity to the headland, to look for a tendency for the cancers to occur on average nearer the headland than expected. Inspection of the data immediately reveals that all ten cases were in the nearest eight

datazones, all but one of which had at least one case. For a formal test of this spatial concentration, a non-parametric conditional test was carried out.

In the course of the calculations for the 10th COMARE Report (2005) entitled "The incidence of childhood cancer around nuclear installations in Great Britain", a considerable amount of work was done on good tests for use on datasets that might in the event involve very small numbers. A test that showed consistently good properties was the "Linear Risk Score" (LRS) distance test, which simply scores each case by the reciprocal of the distance from the putative source of risk and sums these scores for the whole sample. The value of this statistic is then assessed by simulating cases distributed at random.

There are two forms for a test of this kind – the conditional and the unconditional. The former is used to assess the strength of a spatial distribution taking no account of the fact that there may be an overall excess in the area – as is the case with the liver cancers. It is useful for supplying supplementary evidence. The unconditional version integrates the spatial evidence with the overall evidence of raised incidence.

In 10,000 simulations where the 10 cases were re-distributed randomly amongst the datazones (according to size, of course), 328 gave a value of the LRS test statistic greater than that observed in the actual data (13 km^{-1}). This implies a result for this conditional test that is significant at around the 3% level.

The unconditional test samples frequencies in each of the 13 datazones afresh using the Poisson distribution; the total number of cases will usually be much less than 10 of course. Unsurprisingly, therefore, this test gives a much more significant result: in 10,000 simulations only 16 returned a value larger than the observed value, implying an estimated significance level around 0.0016.

4. Conclusions

Examination of the counts in the Dalgety Bay area are quite suggestive of a raised risk of liver cancer, even allowing for multiple testing. Evidence for other tumours is less convincing, but it would be worth further investigation of the lymphomas, both groups of which show an increase in incidence. Spatial analysis of the liver cancer data reinforces the suspicion for this tumour, with a pronounced tendency for the observed cases to be close to the headland. It is recommended that there should be further investigation of these cases and an analysis of earlier data if available.

October 1st 2012