11. LUNG CANCER AND ROASTER EXPOSURE—DESCRIBING THE RISK PATTERN

11.1 Introduction

On the basis of the matched material presented in the previous chapter, it was demonstrated that the work at the roasters was the factor which most probably increased the individual worker's risk of contracting lung cancer. In Fig. 11.1 the lines, symbolising working periods at the roasters, are considerably fewer and shorter for the referents than for the cases. It should be stressed that the number of individuals is the same in both groups.

Even if there was no doubt about the association between roaster work and the occurrence of lung cancer, it is not directly evident how the risk pattern can be described numerically. In this chapter we will demonstrate how elementary standard methods for the analysis of case-referent data are used to estimate the risk for a certain illness as a consequence of a hazardous exposure of some kind. We will indicate what type of conclusions these methods lead to and as a basis we will use the actual data on lung cancer and roaster work given in Table 11.1.

11.2 Non-paired analysis

When cases and referents are matched, it seems most logical to keep the matched pairs together in the statistical analysis. However, as has been pointed out already by Miettinen (1970), the paired nature of the data can be ignored if the correlation within the pairs with regard to the background factor under study is small. As this is the case in this example, we will first make an analysis where the statistical unit is the individual—case or referent—and not the pair.

11.2.1 *The relative risk*

A first, very rough approach is to classify the individuals (cases as well as referents) with regard to whether or not

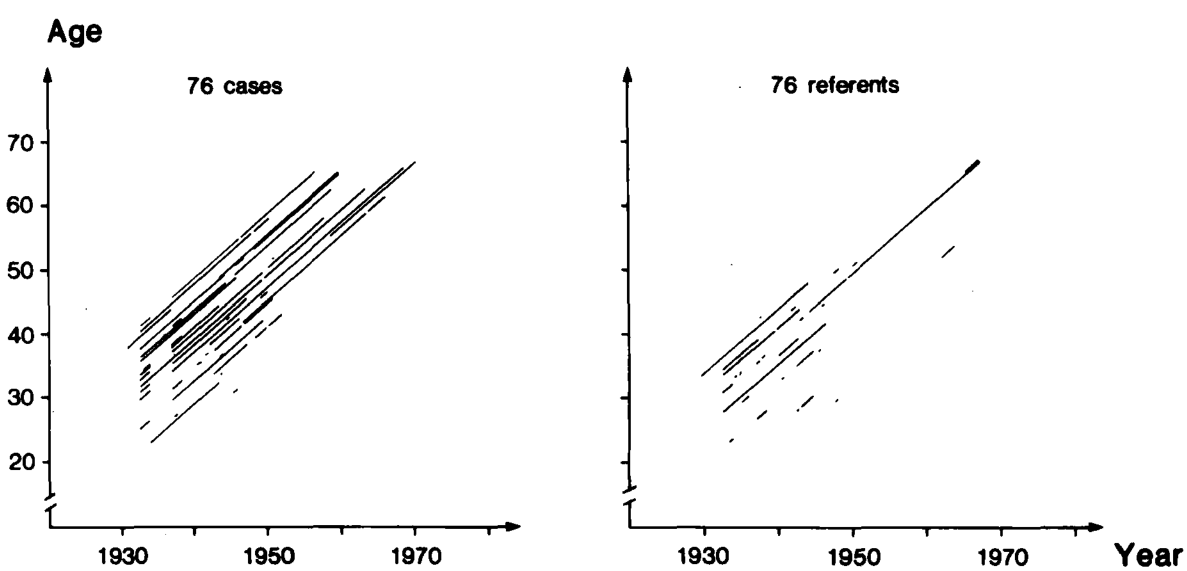

Fig. 11.1. Patterns of exposure at the roaster departments (nos 7–8) among lung cancer cases and matched referents.

Table 11.1

Birth year and exposure time (months) at the roasters (nos. 7–8) for all 76 lung cancer cases and their matched referents. (= case died after 65 years of age)*

Year of birth	Exposure at roasters		Year of birth	Exposure at roasters	
	Case	Referent		Case	Referent
1905*	101	0	1909	0	1
1901	141	26	1910	112	0
1904*	236	167	1901	0	0
1907	6	0	1898*	97	122
1896*	0	0	1910	0	0
1896	301	0	1907	167	33
1902*	399	0	1917	0	0
1905	0	0	1896*	0	0
1909	3	0	1913	0	0
1911	0	0	1911	0	0
1905	17	0	1903*	0	0
1907	0	0	1891*	252	0
1914	0	0	1919	0	0
1900*	0	289	1908*	0	0
1894*	228	2	1925	0	0
1882	3	2	1914	0	0
1900*	0	21	1918	0	4
1907	0	0	1898	27	0
1909	1	0	1914	6	25
1901	0	5	1895	17	0
1903*	50	0	1902*	5	0
1894*	301	0	1900*	370	7
1908	93	0	1899	260	0
1903	0	34	1908	0	3
1898*	0	73	1873	0	0
1910	2	23	1913	0	0
1915	0	0	1896	5	176
1919	0	0	1913	0	0
1903*	152	0	1905*	0	0
1899*	0	9	1900*	1	0
1892	212	0	1896	160	0
1909	0	0	1902*	0	0
1907	0	0	1910	6	23
1910	2	0	1905*	0	11
1906	0	0	1892	73	0
1923	0	0	1909*	0	0
1909	0	0	1909*	0	0
1898	0	0	1900*	1	0

they have ever worked at the roasters. This dichotomization means that we use only the data given in Table 11.2, thus neglecting the detailed information concerning the various lengths of the individual's exposure periods.

As has been shown by Cornfield (1956), further studied by Miettinen (1976) and described in a number of books (Lilienfeld and Lilienfeld 1980, MacMahon and Pugh 1970, Miettinen 1985), it is possible to use the so called odds ratio (OR) as an estimate of the incidence density ratio, sometimes called the relative risk (RR). Hence, the expression

$$RR \approx OR = \frac{A}{C} : \frac{B}{D} = \frac{AD}{BC}$$

which here is $34 \cdot 55/(21 \cdot 42) = 2.12$ and can be interpreted as meaning that the risk of getting lung cancer is roughly twice as large for the exposed compared with the non-exposed.

The value 2.12 is an *estimate* of a postulated 'true' relative risk, RR. It seems reasonable to investigate whether it might only be that the relative risk estimate deviates from unity due to random fluctuations. The null hypothesis H_0:RR = 1 can be tested by means of the ordinary χ^2 test with one degree of freedom, computed as

$$\chi^2 = (AD-BC)^2(T-1)/(N_1 N_0 M_1 M_0)$$

which gives $\chi^2 = 4.83$ corresponding to $p = 0.028$. Since

Table 11.2

Number of individuals among the cases and referents who have either worked (exposed) or not worked (non-exposed) at the roasters

Category	Cases	Referents	Total
Exposed	$A = 34$	$B = 21$	$N_1 = 55$
Non-exposed	$C = 42$	$D = 55$	$N_0 = 97$
Total	$M_1 = 76$	$M_0 = 76$	$T = 152$

$p < 0.05$ we conclude that the result is 'significant on the 5% level', which gives us an indication that the null hypothesis is incorrect.

11.2.2 *Test-based confidence intervals*

The value RR is only a point estimate, revealing nothing about the variability behind the numerical result. Therefore, it might be more elucidating to give an interval estimate. A comfortable way of doing this is to calculate a so called test-based confidence interval. This method was suggested by Miettinen (1976) and, by means of a very simple procedure, it gives approximate confidence limits. In many situations, these approximations are very near the exact confidence limits, which can be quite laborious to calculate. More often, it is the lower test-based confidence limit which is close to the limit obtained through more exact procedures, whereas the upper limit is usually rather too low. For a discussion concerning the actual methods, see works by Fleiss (1979), Finney (1979) and Miettinen (1985).

The basic idea behind the test-based confidence intervals could be described as follows. We test, say, $H_0{:}RR = 1$ by means of the χ^2 test presented above, and we have the estimator $RR = AD/(BC)$. This estimator has a skew distribution and its standard deviation is not known. There are however good reasons to believe that lnRR is approximately normally distributed. Hence, the critical ratio

$$z = (\ln RR - \ln RR_0)/(SE(\ln RR)$$

would be normally distributed with a zero mean and a variance of one if the hypothesis $H_0{:}\ \ln RR = 0$ is true. Thus z^2 corresponds to the χ^2 value in the above-mentioned test of $H_0{:}RR = 1$. A proxy of the standard error of the variable ln RR can therefore be obtained as

$$SE(\ln RR) = \ln RR/\sqrt{\chi^2}$$

and the 95% confidence limits for ln RR as

$$\ln RR \pm 1.96 \ln RR/\sqrt{\chi^2}$$

Taking the antilogarithms and using the data in Table 11.2 we obtain

$$(\underline{RR};\ \overline{RR}) = RR^{(1 \pm z/\sqrt{\chi^2})} = 2.12^{(1 \pm 1.96/\sqrt{4.83})}$$

$$= (1.08;\ 4.16)$$

Table 11.3

Exposure times (in months) at the roasters, for lung cancer cases and referents

Exposure	Cases	Referents	Total
0	$(C =)\ A_0 = 42$	$(D =)\ B_0 = 55$	$N_0 = 97$
1–11	$A_1 = 12$	$B_1 = 9$	$N_1 = 21$
12–23	$A_2 = 2$	$B_2 = 3$	$N_2 = 5$
24–35	$A_3 = 1$	$B_3 = 4$	$N_3 = 5$
36–47	$A_4 = 0$	$B_4 = 0$	$N_4 = 0$
48–59	$A_5 = 1$	$B_5 = 0$	$N_5 = 1$
60–	$A_6 = 18$	$B_6 = 5$	$N_6 = 23$
Total	$M_1 = 76$	$M_0 = 76$	$N = 152$

We used a 95% confidence level ($z = 1.96$) in the above analysis. If instead we had used a 99% confidence level ($z = 2.58$), the interval would have been longer (0.88; 5.12). This latter interval contains the hypothetical value $RR = 1$. It might be worth mentioning that even though the test-based confidence interval gives limits which are approximate, a contradiction between such confidence limits and the corresponding test of significance will never occur.

11.2.3 *Effects of various exposure levels*

If we want to utilize the information on the length of time the individuals spent at the roasters, the data to be used could be presented as in Table 11.3. Here, comparing the display in Table 11.2, we notice that we have $A_0 = C$, $B_0 = D$, $A_1 + A_2 + \ldots + A_6 = A$ and $B_1 + B_2 + \ldots + B_6 = B$.

In the previous analysis, the 'exposed' $(A + B)$ group, indicated by $(0+)$ in Table 11.4 contains a number of individuals $(A_1 + B_1)$ who have been working at the roasters for less than one year. In fact, the majority of them have worked for less than half a year. It seems questionable whether these persons have worked at the roasters long enough to have increased their lung cancer risk. It is probably more informative to compare the non-exposed with those who have been exposed for, say, more than one year $(1+)$. We then obtain the estimator

$$RR_{1+} = (A - A_1) \cdot D/[(B - B_1) \cdot C]$$

which gives the estimator $RR_{1+} = 2.40$. Once again, this is a relative risk estimate, analogous with RR previously calculated from the data in Table 11.2. The corresponding test-based confidence interval is (1.08;5.35).

We can contrast the 'non-exposed' $(C + D)$ with those exposed for more than, say, t years $(A - A_1 - A_2 - \cdots - A_t;\ B - B_1 - B_2 - \cdots - B_t)$, by calculating RR_{t+} for $t = 0, 1, \ldots, 5$ in the same manner as in Table 11.4. It should be pointed out that neither the point estimates, nor the confidence intervals at the various exposure levels are independent of each other, since the different risk estimates are partly based on the same data.

Table 11.4

The relative risk of contracting lung cancer, estimated for various exposure levels

Exposure in years	RR	95% C.L.
0	1.00	—
0+	2.12	1.08; 4.16
1+	2.40	1.08; 5.35
2+	2.91	1.22; 6.92
3+	4.98	1.82; 13.6
4+	4.98	1.92; 13.6
5+	4.71	1.71; 12.9

In Table 11.4 we notice that there appears to be a certain trend in the RR values, making the risk of getting lung cancer approximately 5 times larger for those who have worked at least 3 years at the roasters, as compared with those who have never worked there. It must be pointed out that there are very few individuals in the various groups with exposure times above 24 months. This means that we have no detailed information about the increase in the relative risk RR. The fact that we get the same estimate RR = 4.98 for both t = 3 and t = 4 is due to the circumstance that $A_4 = B_4 = 0$.

11.2.4 *The age factor*

The risk of contracting lung cancer and the length of the exposure period are both related to the individual's age. This circumstance was the reason why the referents were matched with regard to birth year and it was also because of this that we only use the information about the referents up to the time of the death of the corresponding cases. Furthermore, we observe that the roaster exposure might not have been equally hazardous all the time—the work environment has not been the same during all the decades under study.

In the present analysis we are comparing individuals, not pairs. Therefore, the age factor might come in as a *confounder*, in spite of the matching procedure applied when the referents were selected. In order to clarify the role of the age factor, the individuals have been classified into three strata with regard to birth year, as can be seen in Table 11.5.

From Table 11.5 it is obvious that the pattern is not the same within the three strata. We get an estimate RR = 4.69** among those born before 1900, which deviates significantly from unity.

In order to obtain some sort of overall measure of the relative risk, we have to pool the estimates from the various strata. This can be done in a number of different ways.

A maximum-likelihood estimate based on data from all the strata is $RR_{ML} = 2.18$. A weighted overall relative risk, usually giving a value in the neighbourhood of RR_{ML}, was suggested by Mantel and Haenzel (1959) as

$$RR_{MH} = \sum \frac{A_i D_i}{T_i} : \sum \frac{B_i C_i}{T_i}$$

which, for the actual data, gives the estimate $RR_{MH} = 2.16$.

If there is no association between exposure and the occurrence of the illness under study, the proportion of cases is expected to be the same for the non-exposed as for the exposed individuals. If, on the other hand, the influence of the exposure on the risk for the actual illness is strong, we expect the number of cases to be smaller among the non-exposed than among the exposed.

For the three strata together, the total number of unexposed cases is $\Sigma C_i = 6 + 20 + 16 = 42$. Is this amount much smaller than the expected number of unexposed cases ΣC_i, if the proportion of cases among the non exposed had been the same as among the exposed? We can calculate the expected number of unexposed cases as

$$\Sigma C_i' = \Sigma A_i D_i / B_i = 13 \cdot 13/6 + 16 \cdot 26/10 + 5 \cdot 16/5$$
$$= 85.8$$

Table 11.5

Associations between roaster exposure and lung cancer. Individuals stratified with regard to year of birth

Year of birth	Category	Exposure		Relative risk estimate	
		0+	0	RR	95% C.L
-1899	Cases	13	6		
	Referents	6	13	4.69	1.58;13.9
1900-09	Cases	16	20		
	Referents	10	26	2.08	0.78; 5.55
1910-	Cases	5	16		
	Referents	5	16	1.0	—
Total	Cases	34	42		
	Referents	21	55	2.12	1.08; 4.16

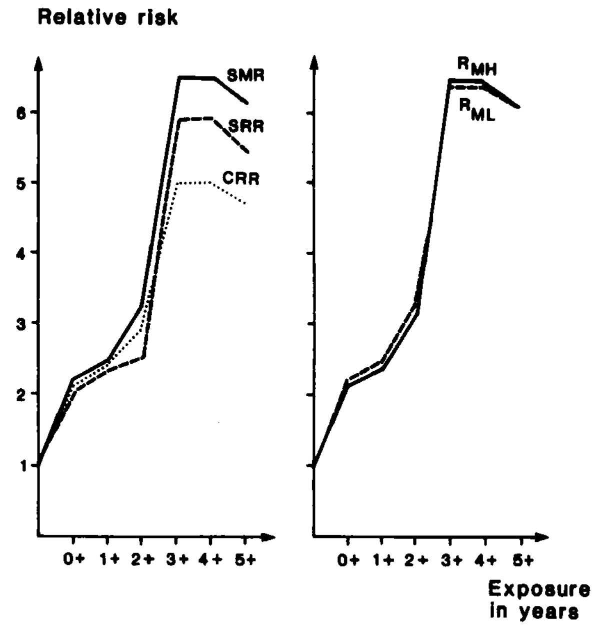

Fig. 11.2. The various suggested relative risk estimates, as functions of roaster-exposure levels.

The ratio $\Sigma C_i'/\Sigma C_i = 85.8/42 = 2.04$ is then the Standardized Rate Ratio (SRR). Since the odds ratio for each stratum is $RR_i = A_iD_i/(B_iC_i)$ $(i = 1,2,3)$, it follows that

$$SRR = \sum C_i RR_i / \sum C_i$$

This means that the SRR is a weighted average of the odds ratios for the different strata and that the weights are proportional to the number of unexposed cases.

In a similar fashion, we can ask whether the number of exposed cases (ΣA_i) is considerably greater than what could be expected if they had had the same proportion of cases and referents as the non-exposed individuals. This means that the group of non-exposed individuals is taken as a standard when the expected number of exposed cases is calculated as

$$\Sigma A_i' = \Sigma B_iC_i/D_i = 6 \cdot 6/13 + 10 \cdot 20/26 + 5 \cdot 16/16$$
$$= 15.46.$$

The ratio $\Sigma A_i/\Sigma A_i' = 34/15.46 = 2.20$ is called the Standardized Mortality Ratio (SMR).

We find, at least on the basis of the actual data, that the various suggested estimators of a summary relative risk give quite similar results. Which one to choose seems rather to be a question of principal than of any practical importance (Fig. 11.2). In a study where various levels of exposure are taken into account, the SRR estimator is preferred, since the weighting system (i.e. the number of unexposed cases) remains the same all the time. This is illustrated in Table 11.6, where the denominator does not change when the SRR's are calculated for various exposure levels. The differences between SRR and SMR, however small, reflect confounding from the age factor, something which SMR does not fully capture. Since the length of exposure is positively correlated with the age, the increasing value of SMR by age is a measure not only of the effects of increasing exposure but also of increasing age.

Table 11.6

A comparison between the direct (SSR) and indirect (SMR) standardization procedure by exposure category. Interval estimates are based on the Mantel-Haenzel point estimate

Exposure in years	SRR	SMR	RR_{MH}	Tested-based C.L.
0+	2.04	2.20	2.15	1.06; 5.67
1+	2.29	2.44	2.35	1.09; 4.25
2+	2.34	2.86	2.79	1.29; 7.59
3+	5.93	6.50	6.52	2.10; 17.81
4+	5.93	6.50	6.52	2.10; 17.81
5+	5.43	6.16	6.16	1.96; 16.98

An overall test of independence between exposure and illness is given by the Mantel-Haenzel χ^2 test with one degree of freedom, where

$$\chi^2_{MH} = \frac{\Sigma A_i - \Sigma(M_{1i}N_{1i}/T_i)}{\Sigma M_{0i}N_{1i}M_{1i}N_{0i}/[T_i^2(T_1 - 1)]}$$

The expressions $M_{1i}N_{1i}/T_i$ $(i = 1,2,3)$ give, for each stratum, the expected frequency of exposed cases, calculated on the assumption of independence between exposure and illness. For the actual data, we get $\chi^2_{MH} = 4.91$*. Thus, the data in Table 11.5 indicate that the hypothesis of a relative risk of one can be rejected on the 5% significance level.

We use the above χ^2_{MH} value when computing a test-based confidence interval for the relative risk as

$$2.16^{(1 \pm 1.96/\sqrt{4.91})} = (1.09; 4.25).$$

In Table 11.6, the 95% confidence limits for the relative risk have been calculated for each exposure level. As was briefly mentioned in connection with Table 11.4, the limits must be interpreted with caution—the set of intervals does not constitute any sort of total confidence region. We stress once more that the same data are partly used in the estimates corresponding to the various exposure levels. Alternatively, a trend test suggested by Mantel (1963) could be used or a regression equation be fitted to the set of SRR-values as recommended by Rothman (1986).

11.2.5 *The truncation of data*

The collection of information for this investigation ended in 1976. This means that the data are censored so that among those born after 1910, the lung cancer cases and their corresponding referents are at most 65 years old and, among those in the middle birth-year stratum (1900–1909), the individuals are at most 75 years old (Fig. 11.3).

Of course, it could be argued that the results from the three birth-year strata are not comparable. For example,

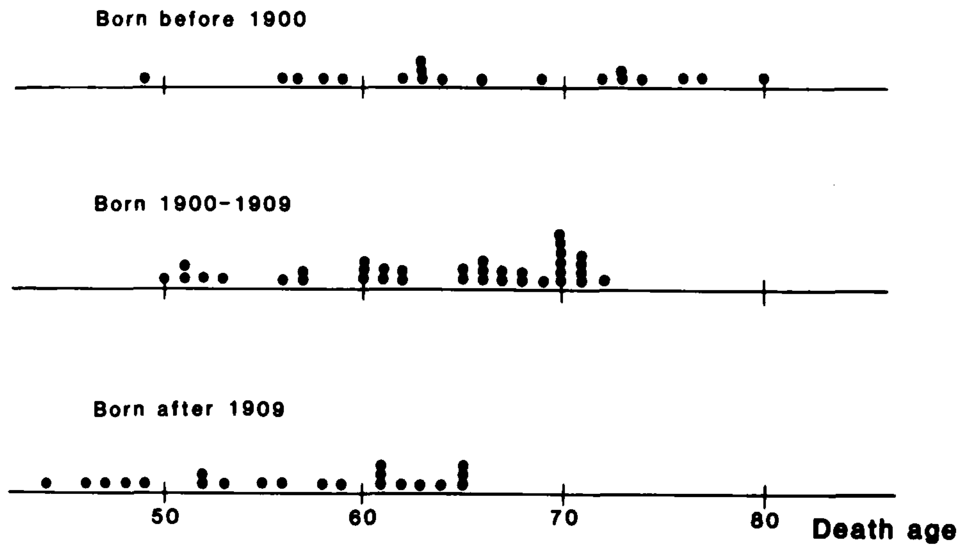

Fig. 11.3. Death age of cases by period of birth.

the hypothesis that all cases of lung cancer specifically caused by roaster work occur after 65 years of age, is not contradictory to the data in Table 11.5. Therefore, as a check, we run separate analyses of all individuals under 65 years of age for the three birth-year strata. This means that we delete from the two older strata all those case-referent pairs older than the individuals in the youngest stratum—the three strata are thus censored in the same way.

Frequencies in Table 11.7 are small in the various cells, but it can be seen that the same pattern is found as in Table 11.5: there is an obvious association between roaster work and the occurrence of lung cancer in the first birth-year stratum but not in the other two strata.

Table 11.7

Associations between roaster exposure and lung cancer before 65 years of age. Stratification with regard to birth year

Year of birth	Category	Exposure 0+	Exposure 0	RR
-1899	Cases	9	2	
	Referents	2	9	20.25
1900-1909	Cases	7	11	
	Referents	5	13	1.65
1910-	Cases	5	16	
	Referents	5	16	1.00

11.2.6 *Conclusions thus far*

Up to this point, based on the individual data and neglecting the paired nature of the material, our analysis has shown that the risk of getting lung cancer is about twice as high for those who have at some time worked at the roasters as for those who have never worked there.

There seems to be a relationship between working period at the roasters and the lung cancer occurrence, so that a person who has worked there for more than three years has a risk of getting lung cancer which is about five times higher than for those who have never worked there.

The strongest association between roaster work and the occurrence of lung cancer is found among the individuals born before 1900.

11.3 Analysis based on the pairs of case and referent

The data at hand consist of 76 pairs and when dichotomizing into 'exposed (0 +)' and 'non-exposed (0)' individuals, the frequencies given in Table 11.8 are obtained.

Table 11.8

Distribution of the matched pairs with regard to whether or not the individuals have ever been exposed to the roasters

Case	Referent 0+	Referent 0	Total
0+	a = 11	b = 23	A = 34
0	c = 10	d = 32	C = 42
Total	B = 21	D = 55	n = 76

11.3.1 *Estimation of the relative risk*

The individuals within each pair should now be compared with each other and it is obvious that there are a + d = 43 pairs, where the case and the corresponding referent have the same classification with regard to exposure. It is only from the other b + c pairs, where the two individuals differ, that we can obtain information about a possible association between lung cancer and the actual exposure. If there were no such associations, it would be expected that the frequencies b and c were equal. Mantel and Haenzel (1959) have shown that the relative risk can be estimated by

$RR_p = b/c = 23/10 = 2.3$

The corresponding test of H_0:RR = 1 is the well-known McNemar test (McNemar, 1947) where

$$\chi^2 = (b - c)^2/(b + c)$$
$$= (23 - 10)^2/(23 + 10) = 5.12^* \text{ 1 d.f}$$

This formula makes it possible to calculate a test-based confidence interval for the relative risk. Here, the 95% confidence limits are 1.12 and 4.7.

Thus the analysis, based on the pairs, gave very similar results to those obtained on the basis of Table 11.2, where the matched pairs were not kept together. The point estimate was $RR_p = 2.3$ instead of RR = 2.12 and the χ^2 value was 5.12 instead of 4.83. Consequently, the confidence interval is somewhat wider when the unit of analysis is the matched pair. In fact, the two point estimators of the relative risk have the following relation (Taube, 1969).

$$RR_p - RR = (RR_p - 1)(ad - bc)/(a + c)(c + d)]$$

showing that they coincide if the relative risk RR = 1 or when ad = bc (which means that there is no correlation within the matched pairs with regard to the exposure variable).

11.3.2 *The age factor, yet again*

The matching procedure aims to make the cases and the referents as similar as possible with regard to one or several background factors—thus 'eliminating' these factors. If this elimination really is achieved, however, it does not necessarily mean that the actual background factor(s) should be neglected in the subsequent analysis.

As in the previous unpaired analysis, we now stratify the units into three birth-year strata and, based on the matched pairs, as is shown in Table 11.9, the analysis produces the same pattern as was found earlier: the strongest association between lung cancer occurrence and roaster exposure is found among those born before 1900; but, in the various strata, the frequencies b and c are so small that no significance is obtained—not even in the first birth-year stratum.

If we then make a separate study of only those pairs where the case died before 65 years of age, the pattern does not change but the frequencies will be still smaller.

11.3.3 *Conclusions thus far*

The analysis based on the pairs reveals the same overall pattern as the previous approach, where the pairs were not kept together. The relative risk estimates based on pairs are generally somewhat larger and the corresponding confidence intervals considerably wider than in the non-paired analysis. The similarity between the results from the two types of analysis of the actual data is due to the fact that the correlation within the pairs was small with regard to exposure. The analysis based upon individuals, not the pairs, is to be preferred for this particular data set.

11.4 Lung cancer risks at all various work sites

In this chapter we have not yet distinguished between the two different types of roaster exposure (nos. 7 and 8) as we did in the previous screening chapter (Chapter 10).

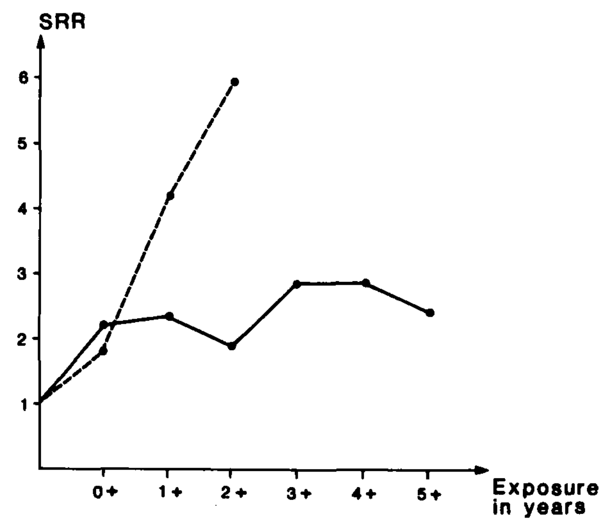

Fig. 11.4. Standardized rate ratios (SRR) by length of exposure at the two roaster departments. - - - Roaster gas purifier (no 8); —— Copper roaster (no 7).

Table 11.9

Distribution of the matched pairs within the different birth-year strata

Category		Frequency	Birth year stratum and exposure categories			Total
Case	Referent		-1899	1900-1909	1910-	
0+	0+	a	4	4	3	11
0+	0	b	9	12	2	23
0	0+	c	2	6	2	10
0	0	d	4	14	14	32
Total		n	19	36	21	76
Relative risk estimate		RR_p	4.5	2.0	1.0	2.3

The reason for this was that, for the methodological presentation here, it was comfortable to have a fairly large proportion of exposed individuals, among both cases and referents. Therefore, we now present such an analysis, treating the two departments separately.

Figure 11.4 shows that the SRR increases dramatically with increasing length of exposure at work site no 8 (the gas purifier). This exposure was not in fact significant when only the 0+ level was considered. The SRR values for the other roaster are well above unity, even if the trend is not equally dramatic.

Interest so far has been focused on the association between lung cancer and work at the roasters. This was natural since it was shown in the previous chapter that such an association did exist. However, the same type of calculations as those just demonstrated could be made for each of the possible exposures (work sites).

Table 11.10 shows the SMRs calculated for the possible association between lung cancer and various exposures, on the basis of the 76 cases and the 76 referents. Several work sites were too small to have a sufficient number of workers exposed for any length of time. Therefore, the SMR-calculations were only performed for the 0+ level.

Table 11.10

Standardized mortality ratios (SMR) for lung cancer among workers ever exposed (0+) at the different work sites, relative to those non-exposed (0)

Work site	SMR	95% C.L.
(1) Ore dressing	1.27	0.49; 3.3
(2) Found works	1.93	0.87; 4.5
(3) Sulphur plant	1.00	—
(4) External works	1.36	0.57; 3.1
(5) Mason	—	—
(6) Conc. delivery	—	—
(7) Copper roaster	2.15	1.04; 4.5
(8) Roaster gas purifier	1.69	0.79; 3.7
(9) Reverb/electric copper furnace	0.60	0.30; 1.4
(10) Converter hall	0.56	0.32; 1.3
(11) Anode furnace	0.80	0.30; 2.1
(12) Coal crusher	—	—
(13) Electrolysis plant	0.60	0.19; 1.9
(14) Cathode furnace/casting	0.51	0.17; 1.6
(15) Nickel smelter	1.60	0.57; 4.1
(16) Precious metals plant	—	—
(17) Arsenic refinery/metal plant	2.12	0.90; 5.1
(18) Arsenic salt/selenium works	0.91	—
(19) Ball foundry	—	—
(20) Sulph acid works	—	—
(21) Lead depts	0.65	0.24; 1.9
(22) Central laboratory	—	—
(23) Machine shop	1.10	0.55; 2.5
(24) Electric shop	4.24	1.00;18.5
(25) Transportation dept	1.42	0.68; 3.1
(26) Building dept	0.53	0.27; 1.13
(27) Various large depts	0.77	0.39; 1.6
(28) Various small depts	1.50	0.42; 5.7

Only the roaster (no. 7) and the electric shop (no. 24) give confidence intervals not containing unity. The electric work shop was the other work site which, in the previous chapter, appeared to have an association with the occurrence of lung cancer.

11.5 More than one referent/case

It is often convenient to have just one referent for each case, especially if the analysis is performed with pairs as the unit of statistical analysis, but it is not self-evident that we should take only one referent for each case. For example, if the number of cases is by necessity very limited—should this circumstance lead to an equally small number of referents? The relatively low additional costs usually motivate selecting two or even three referents for each case.

The 76 referents used for the calculations in this chapter were all selected so that, for each case, the nearest individual fulfilling the specified criteria was selected as a referent (Chapter 10). Since the individuals in the register were ordered according to age, this procedure gave the best possible referents.

If we want to compare the advantage of having one, two or three referents for each case, it seems most meaningful to have a procedure whereby the three referents for each case are equally well matched. We have applied the following procedure: For each case, all individuals fulfilling the criteria (Chapter 10) and born within one year of the birth of the case, were identified. Among these individuals, a simple random sample of 3 individuals was drawn. This means that the first selected referent was not necessarily the best one—there is no special order between the three. Instead they are all equally good, except for random variations.

For testing the possible association between lung cancer and roaster exposure at the various exposure levels, we could use the Mantel-Haenzel statistic. Using the same set of cases—we obtain the same pattern for the actual data, whether we use one, two or three referents. Obviously, the significance would be stronger if we used more than one referent for each case. The set of so-called 'best' referents gives the same results as when we use all three equally matched referents.

The maximum likelihood point estimates of the relative risk for the various exposure levels are approximately the same for the three separate sets of referents and, consequently, the pattern is similar if we use all three referents. When all three are used, however, the confidence limits naturally get closer to each other.