Purpose of this exercise. Nowadays, the calculation of summary measures for stratified data (illustrated in Rothman2002, ch 8) is probably best left to a software program or spreadsheet.[1] But, there are important strategic lessons to be learned from the *structure* of the estimators, both as to when they are/are not appropriate, and because epidemiologists sometimes overlook their structure, and do work that ends up not getting used in the results. Thus, rather than have you program or carry out a large no. of calculations, the exercise is to *inspect* each estimator and see when it is appropriate, and to see the logistical/labour implications – at the design and the analysis stage.

1. Which strata contribute to which summary measures?

i. In his question 8, on pp166-167, Rothman2002 asks you to *imagine* a stratum. Which strata in the *real* example used in the attached Table 1 of the 1959 classic article by Mantel and Haenszel are of this type? Rothman2002 ch 8 & full M-H article in 'Resources/Stratified Data' in 634 site.

ii. What are the contributions of a stratum to the summary *odds* ratio if (a) all of the 'controls' in the stratum are unexposed and all of the cases in the stratum are exposed (b) the converse? Which strata in Mantel and Haenszel's Table 1 are of these types?

iii. We cannot reproduce the ratios (they should be called *IDs*, not *ORs*) in Table 3 in the Ayas et al intern-injuries article (eg 7/31), since they are based on the (self-)matched pairs: see the paragraph beginning "To assess the relationships..." in the last column of page 1057 of the article. For each intern, the data consist of 2 (known) person-time denominators and their corresponding numerators. Thus, both the pooled *difference* (eqn 8-4) and *ratio* (eqn 8-5) measures are calculable from these layouts. The self-paired data consist of over 2000 layouts like 'exhibit i' in the middle of p147 Rothman2002 Ch 8, i.e., i goes from 1 to 2000-something.

Question: When you apply eqns 8-4 (difference) and 8-5 (ratio) to these matched sets, which ones do and do not affect each summary measure?[2]

iv. Imagine you were a reviewer for the NEJM article by Ayas et al. What would you have said in your review about the authors' use of 'OR'?

2. How well does Woolf's summary odds ratio measure[3] work with the lung cancer data (cf M-H) and bladder cancer data (cf Miettinen)?

i. The frequencies in the examples in Woolf's paper are *quite* large. How well/poorly does Woolf's method work if strata have sparse data? For example, what would happen if this method were used to calculate a summary odds ratio and associated 95% CI from the strata in the M-H article?[4]

3. The etiologic study Inspect the *point estimates* whose variability[5] is described by eqns 7-5 and 7-6. Rewrite the $OR = \widehat{\widehat{IDR}} = ad/bc$ of page 139 so that it has the same form as the $\widehat{IDR} = (a/PT_1) \div (b/PT_0)$ on page 137. Explain why PT_1 and PT_0 do not appear in eqn 7-5, whereas their counterparts c and d do in eqn 7-6.

4. How big a denominator ("""control""") series?

i. In the studies cited by Miettinen and by Woolf, restrict attention to estimation of the IDR in the first row (age 50-55, London) respectively. Suppose the numbers of new cases (bladder cancer, peptic ulcer) were limited to those listed, but that the size of the denominator series could have been increased/decreased. The CI for IDR is based on Rothman2002 7-6.

Had the denominator series been scaled up/down, the 'denominator' numbers would not stay in the exact same ratio[6], but say for the moment that they would: e.g., had the numbers interviewed been 260, the breakdown would be somewhere near 220:40. Plot (on y axis) the SEs based on numerator:denominator series of 25:13, 25:26, 25:52, 25:104, up to 25:260, then 25:500 and 25:2500 against on x-axis the control/case ratio. Comment.[7]

ii. Does the same issue apply with the size of the denominator series for London – and especially – Newcastle? (mind you, if the denominator series is from the blood bank, there is less of a budget issue, since the blood-typing work would have been done even if there were no study. The same issue applies when a denominator series is extracted from an administrative database, where it doesn't cost much more for the agency to extract from 1000 electronic records as it does 100, but there is the extra work for the researcher – cleaning a greater numbers of records.

[1] c634 site, Resources for Stratified Data

[2] For one of the two measures, the data supplied to us by Dr Ayas are in a spreadsheet under Resources for Stratified Data.

[3] Woolf's 1955 paper is also available under Resources for Stratified Data.

[4] The M-H estimator of a common ratio works well for strata of all sizes, from ones like in the Woolf examples, down to strata consisting of matched pairs. The "Robins-Greenland-Breslow" formula (the last one in Rothman2002 Table 8-4) for $Var[\log OR_{MH}]$ was developed in 1986. It is not a lot of work if you program it (once) in a spreadsheet (see Resources), but is tedious by calculator – a lot of $\times \div +$ and $-$. The third competitor – the *test-based* CI, descibed in JH's "Notes on stratified data" – is the simplest of the three, especially if one has already calculated the null chi-squared test statistic.

[5] In the log scale.

[6] After all, the aim of a larger denominator series is to obtain a more reproducible estimate of the exposure distribution in the source.

[7] jh discusses the effect of the denominator:case series ratio in his 607 notes ch 8.2, p 8.

TABLE 1.—*Illustrative computations for chi square and for summary measures of relative risk (R, R_1, R_2, R_3, and R_4) relating to the association of epidermoid and undifferentiated pulmonary carcinoma in women with smoking history*

Group		Epidermoid-undifferentiated pulmonary carcinoma			Controls			Cases and controls			Derivative computations								
		1+ Pack cigarettes daily	Nonsmokers	Total	1+ Pack cigarettes daily	Nonsmokers	Total	1+ Pack cigarettes daily	Nonsmokers	Total	$\frac{AD}{T}$ $\frac{(1)(5)}{(9)}$	$\frac{BC}{T}$ $\frac{(2)(4)}{(9)}$	$\frac{E(A)}{(3)(7)}$ $\frac{}{(9)}$	$\frac{E(D)}{(6)(8)}$ $\frac{}{(9)}$	$\frac{V(A)}{(12)(13)}$ $\frac{}{(9)-1.0}$	$\frac{N_1C}{N_2}$ $\frac{(3)(4)}{(6)}$	$\frac{N_1D}{N_2}$ $\frac{(3)(5)}{(6)}$	$\frac{N_2A}{N_1}$ $\frac{(1)(6)}{(3)}$	$\frac{N_2B}{N_1}$ $\frac{(2)(6)}{(3)}$
		A (1)	B (2)	N_1 (3)	C (4)	D (5)	N_2 (6)	M_1 (7)	M_2 (8)	T (9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Housewives	under age 45	0	2	2	0	7	7	0	9	9	0	0	0	7.000	0	0	2.000	0	7.000
	45–54	2	5	7	1	24	25	3	29	32	1.500	0.156	0.656	22.656	0.480	0.280	6.720	7.143	17.857
	55–64	3	6	9	0	49	49	3	55	58	2.534	0	0.466	46.466	0.380	0	9.000	16.333	32.667
	65 and over	0	11	11	0	42	42	0	53	53	0	0	0	42.000	0	0	11.000	0	42.000
White-collar workers	under age 45	3	0	3	2	6	8	5	6	11	1.636	0	1.364	4.364	0.595	0.750	2.250	8.000	0
	45–54	2	2	4	2	18	20	4	20	24	1.500	0.167	0.667	16.667	0.483	0.400	3.600	10.000	10.000
	55–64	2	4	6	2	23	25	4	27	31	1.484	0.258	0.774	21.774	0.562	0.480	5.520	8.333	16.667
	65 and over	0	6	6	1	11	12	1	17	18	0	0.333	0.333	11.333	0.222	0.500	5.500	0	12.000
Other occupations	under age 45	1	0	1	3	10	13	4	10	14	0.714	0	0.286	9.286	0.204	0.231	.769	13.000	0
	45–54	4	1	5	1	12	13	5	13	18	2.667	0.056	1.389	9.389	0.767	0.385	4.615	10.400	2.600
	55–64	0	6	6	1	19	20	1	25	26	0	0.231	0.231	19.231	0.178	0.300	5.700	0	20.000
	65 and over	1	3	4	0	15	15	1	18	19	0.790	0	0.211	14.211	0.166	0	4.000	3.750	11.250
Total		18	46	64	13	236	249	31	282	313	12.825	1.201	6.375	224.375	4.036	3.325	60.675	76.960	172.040

Checks: Total discrepancy, Y, $= \Sigma A - \Sigma E(A) = \Sigma(1) - \Sigma(12) = 11.625$
$= \Sigma D - \Sigma E(D) = \Sigma(5) - \Sigma(13) = 11.625$
$= \Sigma(AD/T) - \Sigma(BC/T) = \Sigma(10) - \Sigma(11) = 11.625$

$\Sigma(15) + \Sigma(16) = 64.000$; $\Sigma(3) = 64$
$\Sigma(17) + \Sigma(18) = 249.000$; $\Sigma(6) = 249$

Derivative computations: $\Sigma E(B) = \Sigma(2) + Y = 57.625$
$\Sigma E(C) = \Sigma(4) + Y = 24.625$
$\Sigma(AT/N_1) = \Sigma(1) + \Sigma(17) = 94.960$
$\Sigma(BT/N_1) = \Sigma(2) + \Sigma(18) = 218.040$
$\Sigma(CT/N_2) = \Sigma(4) + \Sigma(15) = 16.325$
$\Sigma(DT/N_2) = \Sigma(5) + \Sigma(16) = 296.675$

Chi-square: $X^2 = (|\text{discrepancy}| - 0.5)^2/\Sigma V(A) = (|Y| - 0.5)^2/\Sigma(14) = 30.66$

Relative risk: $R = \Sigma(AD/T)/\Sigma(BC/T) = \Sigma(10)/\Sigma(11) = 10.68$

R_1 { crude relative risk, $r = \Sigma A\Sigma D/\Sigma B\Sigma C = \Sigma(1)\Sigma(5)/\Sigma(2)\Sigma(4) = 7.10$
adjustment factor, $f = \Sigma E(A)\Sigma E(D)/\Sigma E(B)\Sigma E(C) = \Sigma(12\Sigma(13)/\Sigma E(B)\Sigma E(C = 1.0081$
$R_1 = rf = 7.05$

$R_2 = \Sigma A\Sigma(N_1D/N_2)/\Sigma B\Sigma(N_1C/N_2) = \Sigma(1)\Sigma(16)/\Sigma(2)\Sigma(15) = 7.14$

$R_3 = \Sigma(N_2A/N_1)\Sigma D/\Sigma(N_2B/N_1)\Sigma C = \Sigma(5)\Sigma(17)/\Sigma(4)\Sigma(18) = 8.12$

$R_4 = \Sigma(AT/N_1)\Sigma(DT/N_2)/\Sigma(BT/N_1)\Sigma(CT/N_2) = 7.91$

Note: Figures shown are rounded from those actually calculated and consequently are not fully consistent. Column totals and figures shown do not necessarily agree.

ON ESTIMATING THE RELATION BETWEEN BLOOD GROUP AND DISEASE

By BARNET WOOLF

Department of Animal Genetics, University of Edinburgh

Following the demonstration of a significant excess of blood group A in patients with cancer of the stomach (Aird, Bentall & Roberts, 1953) and of group O in sufferers from peptic ulcer (Aird, Bentall, Mehigan & Roberts, 1954) and from toxaemia of pregnancy (Pike & Dickins, 1954) it seems certain that many more studies will be made on the relation between blood groups and disease. It is therefore important that the best possible statistical methods should be used. The procedure recommended by Aird *et al.* (1954) is very efficient, but it is open to criticism on one rather important point. These workers take as criterion the difference in proportion of a given blood group in the disease and the control series. Denote the two blood types α and β. Suppose the disease series contains h patients of type α and k of type β, where $h+k=n$, and the control series has H of type α and K of type β, where $H+K=N$. Aird and associates calculate $d=h/n-H/N$. This is tested for significance against its sampling variance, combined with estimates from other bodies of data to give a weighted mean estimate, and compared with these other estimates in tests for heterogeneity.

Unfortunately, d will differ from one community to another even when the specific attack rate within any given blood group stays constant. This can be shown by a simple example. Consider a community of 10,000 people in which H and K are each 5000. Then if $h=100$ and $k=50$, $d=100/150-0{\cdot}5$, or $0{\cdot}1667$. Now consider another community in which H is 9000 and K is 1000. In this case $h=180$ and $k=10$, so $d=180/190-0{\cdot}9$, or $0{\cdot}0474$. Even when the essential biological conditions are identical, differences in blood-group frequencies in the population will introduce spurious heterogeneity. This kind of artefact is avoided if one works with incidence rates in the various blood groups. The data usually do not permit calculation of absolute rates, nor are they needed. What is wanted and readily obtained is an estimate of the ratio of one rate to another. The incidence in group α will be $h/H \times$ some constant, and that in group β will be $k/K \times$ the same constant. If the ratio is taken as x to 1, an estimate of x will be hK/Hk, and it may readily be shown that this is the maximum-likelihood estimate. The use of x is recommended instead of d as a criterion of differential incidence of disease in relation to blood group.

In all statistical computations it is best to transform x into its logarithm. This avoids difficulties due to asymmetry. If comparison of α with β gives $x=2$ say, comparison of β with α will give $x=\frac{1}{2}$; but $\log x$ will retain its numerical value, merely changing in sign. Moreover, the sampling variance of $\log x$ is a very simple expression free of 'nuisance parameters'. This is especially true if one transforms into $y=\log_e x$. If V is the sampling variance of y, then

$$V=1/h+1/k+1/H+1/K,$$

and w, the weight of y, is of course $1/V$. If the attack rate is the same for both blood types the expected value of y will be 0, so the null hypothesis is not to be rejected unless y differs significantly from zero. This is tested by χ^2 will be y^2/V or wy^2 for one degree of freedom. Combination of data from different communities proceeds as described by Aird *et al.* (1954). The weighted mean, Y,

is $\Sigma wy/\Sigma w$, and its antilogarithm, X, is taken as the combined estimate of x. Significance of Y is tested by $\chi^2=(\Sigma wy)^2/\Sigma w$ or $Y^2\Sigma w$ for one degree of freedom. Heterogeneity is tested by $\chi^2=\Sigma wy^2-(\Sigma wy)^2/\Sigma w$ or $\Sigma wy^2-Y^2\Sigma w$ with degrees of freedom one less than the number of sets of data combined. The standard deviation of y is $V^{\frac{1}{2}}$, and the approximate fiducial limits at the 95% point are $y\pm 1{\cdot}96V^{\frac{1}{2}}$. Provided there is no significant heterogeneity the standard deviation of Y is $1/(\Sigma w)^{\frac{1}{2}}$ and the 95% fiducial limits are approximately $Y\pm 1{\cdot}96/(\Sigma w)^{\frac{1}{2}}$. By taking antilogarithms these can be transformed into fiducial limits for x or X. This is a 'large-sample' treatment, and the formulae cease to be applicable if any of the observed frequencies is small.

Table 1. *Calculation of combined estimate of incidence ratio of peptic ulcer in groups O and A*

City	Peptic ulcer		Control		$x=\frac{hK}{Hk}$	$y=\log_e x$	$w=\frac{1}{\frac{1}{h}+\frac{1}{k}+\frac{1}{H}+\frac{1}{K}}$	$wy^2=\chi^2$
	Group O (h)	Group A (k)	Group O (H)	Group A (K)				
London	911	579	4578	4219	1·4500	0·3716	304·9	42·11
Manchester	361	246	4532	3775	1·2224	0·2008	136·6	5·50
Newcastle	396	219	6598	5261	1·4418	0·3659	134·5	18·01
					$\Sigma wy=189{\cdot}94$		576·0	65·62

$Y=\Sigma wy/\Sigma w=0{\cdot}3289$.
$Y^2\Sigma w=62{\cdot}63$.
S.D. of $Y=(\Sigma w)^{-\frac{1}{2}}=0{\cdot}0417$.
95% fiducial limits of $Y=0{\cdot}2472$–$0{\cdot}4106$.
$X=$ antilog $Y=1{\cdot}39$.
95% fiducial limits of $X=1{\cdot}28$–$1{\cdot}51$.

χ^2 analysis

	D.F.	
Y	1	62·63
Heterogeneity	2	2·99
Total	3	65·62

Table 2. *Incidence ratios of some diseases in relation to blood group*

Disease	Comparison	X or x	95% fiducial limits	Reference
Cancer of stomach	Group A with group O	1·22	1·12–1·32	Aird *et al.* (1953)
Peptic ulcer	Group O with group A	1·39	1·28–1·51	Aird *et al.* (1954)
Toxaemia of pregnancy	Group O with all others	1·38	1·15–1·66	Pike & Dickins (1954)

Table 1 shows the calculations for comparing incidence of peptic ulcer in groups O and A using combined data from London, Manchester and Newcastle. This is the example worked out by their method by Aird *et al.* (1954, Table VII). The χ^2 values, 62·63 for significance and 2·99 for heterogeneity, agree closely with 66·21 and 3·01 found by the d method. Similarly, combined data from six centres on cancer of the stomach in groups O and A gave χ^2 values by the x method of 21·28 (1 D.F.) for significance and 2·63 (5 D.F.) for heterogeneity, against 21·49 and 2·69 by the d method. These close concordances are to be expected, since the observations all come from cities in England with very similar population blood-group frequencies. If data from different ethnic groups were combined, the d method would in general be expected to return unduly high χ^2 figures for heterogeneity. This might be of some biological or medical importance as

tending to be confounded with possible genuine heterogeneity arising either from environmental factors or from differential attack rates for the diverse genotypes that may go to make up a single blood group. Even when heterogeneity is not an issue, x is preferable to d because it has a direct medical meaning. Table 2 gives some estimated incidence ratios of diseases in relation to blood group, together with fiducial limits. A blood-group difference appears able to increase the risk of disease by as much as 39 %.

REFERENCES

AIRD, I., BENTALL, H. H., MEHIGAN, J. A. & ROBERTS, J. A. F. (1954). The blood groups in relation to peptic ulceration and carcinoma of colon, rectum, breast and bronchus. *Brit. Med. J.* **2**, 315.

AIRD, I., BENTALL, H. H. & ROBERTS, J. A. F. (1953). The relationship beteeen cancer of stomach and the *ABO* blood groups. *Brit. Med. J.* **1**, 799.

PIKE, L. A. & DICKINS, A. M. (1954). *ABO* blood groups and toxaemia of pregnancy. *Brit. Med. J.* **2**, 321.