-1a- events in unstratified person-time - Say, as in Example 3.10 of Breslow and Day Vol II, that the data are from a single stratum in which $O_0 = 5$ and $O_1 = 14$, whereas $PT_0 = 7300$ and $PT_1 = 5500$, and interest is in the rate ratio parameter $\psi = \lambda_1/\lambda_0$.

We showed in bios601 that $O_1 \mid O_+ \sim Binomial(O_+, \pi)$, where

$$\pi = \psi \times PT_1/(\psi \times PT_1 + PT_0).$$

i. From this, using 1st principles, derive the ML estimator, $\hat{\psi}_{ML-condn'l}$, of ψ.

ii. Obtain $\hat{\psi}_{ML-condn'l}$ from a generalized linear model. *Hint*: write $\text{logit}(\pi)$ as a function of ψ.

-1b- events in stratified person-time - See section (c) of section 3.6 of Breslow and Day Vol II. There they say that the ML estimation of the rate ratio ψ from stratified PT data requires iterative calculations, so let's iterate...

We will use Example 3.11, with data, shown in Table 3.14, page 111, from $J = 13$ age-period strata. Again interest is in the rate ratio parameter $\psi = \lambda_{j1}/\lambda_{j0}$, assumed (for now) to be constant over the J strata.

Thus, for each of the J strata, $O_{j1} \mid D_j \sim Binomial(D_j, \pi_j)$, where

$$\pi_j = \psi \times PT_{j1}/(\psi \times PT_{j1} + PT_{j0}).$$

Note the switch of notation, from O_{j+} to D_j.

i. Derive the ML estimating equation (3.15) for $\hat{\psi}_{condn'l}$, by obtaining the expression for $d \log L/d\psi$ and setting it to zero.

ii. Use the Newton-Raphson iterative method to find the root of the $d \log L/d\psi$ function, ie

$$\hat{\psi}^{(k+1)} = \hat{\psi}^{(k)} + \left.\frac{d \log L/d\psi}{d^2 \log L/d\psi^2}\right|_{\hat{\psi}^{(k)}} = \hat{\psi}^{(k)} + \left.\frac{\Sigma_j d \log L_j/d\psi}{\Sigma_j d^2 \log L_j/d\psi^2}\right|_{\hat{\psi}^{(k)}}.$$

iii. How does the iteration change if we rewrite the Likelihood, and thus the log Likelihood, in terms of β, where $\psi = \exp(\beta)$?

iv. Obtain $\hat{\psi}_{condn'l}$ from a generalized linear model (Binomial) fitted to the 13 binomial observations. The stratified data are available in the BIOS602 website under the Resources link. Note that one can specify Binomial (rather than Bernoulli) data by using as 'y' a matrix with 2 columns: the numbers positive and negative, i.e. `glm(cbind('# +ve' vector,'#no. -ve' vector)` $\sim \ldots$, `family=binomial, ...)`.

v. Obtain $\hat{\psi}_{uncondn'l}$ from a generalized linear model (Poisson, 14 parameters) fitted to the $(j = 1, \ldots, 13) \times (i = 0, 1) = 26$ observations $\{O_{ji}, PT_{ji}\}$.

Are your estimates in agreement with Breslow and Day's statement (lines 5-6, page 109) that under the Poisson model, $\hat{\psi}_{condn'l} = \hat{\psi}_{uncondn'l}$?
Note B&D's comment that the same will **not** be true for conditional vs. unconditional estimation of a common rate ratio **when the PT's are estimated** from J stratified denominator ('control') series, particularly if the strata are sparse.

-2a- single (unstratified) case series, and denominator series[1]

Say, as in section 4.2 Breslow and Day Vol I, that the data are from a single stratum in which $O_0 =$ 'c' $= 3$ and $O_1 =$ 'a' $= 2$, whereas $\widehat{PT_0 : PT_1} = 1 : 1$, based on a denominator series of 2 person moments, classified into 'd' $= 1$ unexposed person-moment, and 'c' $= 1$ exposed person-moment [in notation of equation 4.1]. As before, interest is in the rate ratio parameter $\psi = \lambda_1/\lambda_0$.

Let us adopt the notation for Design 2 in the 2-page handout from course epib-634[2], namely c_1 and c_0 for exposed and non-exposed *c*ases, and d_1 and d_0 for the numbers of histories of exposure and non-exposure in those persons forming the *d*enominator series.

First, as above, we have explicit statistical models for the 2 numerators (i=1 exposed, i=0 not exposed):

$$c_i \sim Poisson(\mu_i = \lambda_i \times PT_i); \quad \psi = \lambda_1/\lambda_0.$$

Likewise, because the denominator series was formed by simple random sampling of the base, we have an explicit statistical model for the $d_1 : d_0$ split:

$$d_1 \mid (d_1 + d_0 = d) \sim Binomial(d, \pi' = PT_1/(PT_1 + PT_0)).$$

(We don't necessarily have to, and will see below what else we could do) but ***if*** we condition on the sum, c, of the 2 numerators c_1 and c_0, we eliminate the separate parameters λ_1 and λ_0 and are left with the parameter ψ and the ratio of the two unknown constants PT_1 and PT_0, i.e.,

$$c_1 \mid c \sim Binomial(c, \pi), \text{ where } \pi = \psi \times PT_1/(\psi \times PT_1 + PT_0).$$

[1] denonimator series is a sample of the person moments, and serves as an estimate, albeit containing sampling error, of the $PT_1 : PT_0$ ratio. Many, unfortunately, cling to the confusing term '*control*' series, and refer to the people identified at the sampled person moments as 'controls.'

[2] See Notes, March 18: Estimation of IDR: ID in index (1) vs. ID in reference (0) category: 2 versions of the etiologic study.

We also showed, back in bios601 (cf. lecture notes on 2 proportions[3], section 5.3), that if one has two binomial observations, $y \sim Binomial(n, \pi)$ and $y' \sim Binomial(n', \pi')$, then the distribution of y conditional on their sum $y+y'$, is non-central hypergeometric with parameter $\omega = \{\pi/(1-\pi)\} \div \{\pi'/(1-\pi')\}$. In our notation, then, this means that the distribution of c_1, conditional on the sum $c_1 + d_1$, and on the marginal totals $c_1 + c_0$ and $d_1 + d_0$ already fixed by agreement and by design, is non-central hypergeometric with parameter

$$\{\pi/(1-\pi)\} \div \{\pi'/(1-\pi')\} = \frac{\psi \times PT_1}{PT_0} \div \frac{PT_1}{PT_0} = \psi = \lambda_1/\lambda_0.$$

i. Derive the ML estimating equation for $\hat{\psi}_{condn'l}$, by obtaining the expression for $d \log L/d\psi$ and setting it to zero. This is estimating equation 4.5 in Breslow and Day, Volume I (Analysis of Case-Control Studies – chapter in Resources).

ii. One obvious way, especially if the Likelihood involves higher order polynomials, to obtain $\hat{\psi}_{condn'l}$, is by trial and error, i.e., by increasing or decreasing $\hat{\psi}_{condn'l}$, until the fitted frequency in the a cell matches the observed frequency (by the way, it doesn't matter whether one tracks the a, b, c or d frequency, since, with all four marginal totals fixed, there is only one free entry). But this way does not tell us anything about the reliability of the estimate.

Thus, use the Newton-Raphson iterative method to find the root of the $d \log L/d\psi$ function, ie

$$\hat{\psi}^{(k+1)} = \hat{\psi}^{(k)} + \frac{d \log L/d\psi}{d^2 \log L/d\psi^2}\bigg|_{\hat{\psi}^{(k)}} = \hat{\psi}^{(k)} + \frac{\Sigma_j d \log L_j/d\psi}{\Sigma_j d^2 \log L_j/d\psi^2}\bigg|_{\hat{\psi}^{(k)}}.$$

As a by-product you will have the *information*[4], $-\{d^2 \log L/d\psi^2\}$, the inverse of which can be used as the variance for $\hat{\psi}$.

-2b- several (stratified) case series and corresponding denominator series

See the example on page 137 of Breslow and Day Vol I, and repeated in Table 4.3 page 145. The calculations for all of the other estimation methods are shown in their entirety, but not those for the iterative ML methods (both $\hat{\psi}_{uncondn'l}$ and $\hat{\psi}_{condn'l}$), so let's iterate... Because there is a considerable amount of data (96 exposed cases in all), the example does not show the importance of taking the conditional approach – we would need to take an example from Chapter 7 of B&D Volume I in order to fully appreciate this. However, since table 4.3 also shows all of the other estimators, we will use it to show the ML approaches as well. Data are available in Resources.

i. Derive the ML estimating equation (4.25) for $\hat{\psi}_{condn'l}$, by obtaining the expression for $d \log L/d\psi$ and setting it to zero.

ii. Use the Newton-Raphson iterative method to find the root of the $d \log L/d\psi$ function, ie

$$\hat{\psi}^{(k+1)} = \hat{\psi}^{(k)} + \frac{d \log L/d\psi}{d^2 \log L/d\psi^2}\bigg|_{\hat{\psi}^{(k)}} = \hat{\psi}^{(k)} + \frac{\Sigma_j d \log L_j/d\psi}{\Sigma_j d^2 \log L_j/d\psi^2}\bigg|_{\hat{\psi}^{(k)}}.$$

iii. How does the iteration change if we rewrite the Likelihood, and thus the log Likelihood, in terms of β, where $\psi = \exp(\beta)$? See Appendix.

iv. Calculate $\mathrm{Var}[\hat{\psi}]$, using the inverse of the information, $-\{d^2 \log L/d\psi^2\}$,

v. Obtain $\hat{\psi}_{uncondn'l}$ from a generalized linear model (logistic) fitted to the $(j = 1, \ldots, 6) \times (i = 0, 1) = 12$ Binomial observations $\{c_{ji}/[c_{ji} + d_{ji}]\}$.

Breslow and Day (lines 5-6, page 109, Volume II) tell us that under the Poisson model, with ***known*** PT denominators, $\hat{\psi}_{condn'l} = \hat{\psi}_{uncondn'l}$. As is clear from the estimates shown in line 4 of page 144 Volume I, the same is **not** true for conditional vs. unconditional estimation of a common rate ratio **when the PT's are estimated** from J stratified denominator ('control') series, particularly if the strata are sparse.[5] We will see more extreme examples in Chapter 7.

-2c- several risksets in a historical cohort[6] study –

See example 5.1 beginning on page 187 of Breslow and Day Vol II. The time axis is age, and the conditional Likelihood that Cox set up for each riskset[7] means that the Likelihood is exactly the same as the one that leads to $\hat{\psi}_{condn'l}$ in the data in Volume I, Table 4.3.

[3] Section 3 deals with test-based CI's, a topic we did not spent much time on.

[4] a scalar in this example, a matrix when there are ≥ 2 parameters.

[5] cf. Mantel, N. and Hankey, W. (1975). The odds ratio of a 2×2 contingency table. *American Statistician* 29, 143–145, for a nice (single-stratum) proof of the fact that, in a single table representing the data from a case series coupled with estimated denominators (a sample of the base), $\psi_{condn'l}$ is closer to the null than $\hat{\psi}_{uncondn'l}$ is. The same is true for stratified data.

[6] It might be better to view this as an open population, since subjects move from the unexposed to the exposed category during the followup.

[7] Technically, when we have 'ties', i.e., 2 or more events at the same time (age, here), there are a few ways to set up the Likelihood.

i. Use the same computer code you used in 2b to obtain the $\hat{\psi} = 1.80$ that Breslow and Day report on page 189, in the paragraph beginning "The full partial likelihood..." [*The data are available under Resources*].

ii. How easy would it be to modify your code so that it could fit a model in which the hazard ratio (or what Breslow and Day loosely call the relative risk) for oestrogen use "varied with age"? the

-2d- 17 leukemia risksets in the Woburn study[8]

i. Refer to the data in Table 2, available electronically, along with some R code, under Resources. Analyze exposure to wells G and H as a binary variable, 0 vs. >0. You can reconstruct the numbers and unexposed subjects from the percentages in the last column, and the numbers[9] in the risksets.

From the value of the incidence rate ratio (IRR), or of $\beta = \log(IRR)$, that maximizes Cox's Likelihood. See Figure 5 in part II of JH's Oct 5, 2004 draft on survival analysis, risk sets, .. a unified view. You might be able to use the same computer code you used in 2b to obtain this estimate. See also the ML estimation approaches, via R, in the 'Woburn leukemia data' in the Resources for the course.

ii. Can you trick a regular (unconditional) logistic program (e.g., `glm` in `R`) into fitting this incidence rate ratio (IRR)?

iii. Note that in each riskset of the Woburn study, just as in the Hutchinson et al study of hormones and breast cancer (2c above, Table 5.2 of B & D II), the number of 'non-cases' is >> the number of cases. Given this, we might expect that it would not matter (i.e, the variance would not be very different) if we treated the underlying denominators as *known* rather than as *estimated from a sample of the base*. Re-run the analyses in (i) treating the composition of each risk set as fixed rather than random, i.e., use the same estimation method as in 1b. (ii) sampling say 10 and 25 at random from each riskset. Repeat a few times with different riskset samples.

iv. Use this example, with just 1 case per riskset, and the single[10] binary exposure ('z', in Cox's notation), to examine carefully the 'anatomy' of the estimating equation and the detailed iteration steps in arriving at $\hat{\beta} = \log(HR)$. To do so

(a) Derive the score for each riskset, i.e., fill in the 'algebra' that allowed Cox to get from equation (12) to equation (14) in his 1972 paper.

(b) Derive the Information contribution from each riskset, i.e., fill in the 'algebra' that allowed Cox to write equation (16).

(c) Obtain the ML estimate of $\beta = \log IRR$ and its variance. As Cox suggested in the last paragraph of page 191, do so 'by iterative use of (14) and (16) in the usual way'. From these, form a 95% CI for the incidence rate ratio.

What changes if we treat exposure as a continuous variable? Are there sufficient details in Table 2 to allow you to estimate the corresponding parameter? If there are, do so; if not, explain.

-2e Reducing the crying of babies [11]

Refer to the paper "The role of Stimulation in the Delay of Onset of Crying in the Newborn Infant" by T Gordon and B.M. Foss in the Quarterly J of Experimental Psychol., 18, 79-81. 1966. The Gordon and Foss article, the data, and Cox's 1966 data-analysis, are on the c626 website under Datasets (towards bottom of page).

i. Treat each day as a stratum. Set up a model containing an odds ratio ψ, assumed (for now) common over days. Estimate ψ using (a) an unconditional (b) a conditional, logistic regression approach. Comment.

ii. Set up the problem as a 'survival analysis' ['time to event'] one, where say we know the times at which individual babies cried. Use each day as a separate stratum.

iii. Do you need to know the exact times when babies started to cry, or is the order in why they cried sufficient? Explain.

iv. How might you handle what for now look like unrankable (tied) pairs of observations?

[8] Lagakos SW, Wessen BJ, and Zelen M. Analysis of Contaminated Well Water and Health Effects in Woburn, Massuchusetts, JASA 1986, pages 583-596 + Discussion. JASA paper is on jh's c626 website
See also `http://www.geog.ubc.ca/courses/geog471/notes/health/grabber/index.html`

[9] JH is often asked if the riskset includes the cases; the answer is yes – the riskset is the set of *candidates* for the event of interest.

[10] so that the information and variance covariance matrices are of dimension 1×1.

[11] These data are considered in the early editions of Cox's book on the Analysis of Binary data, and in the paper (to be found under Resources) A Simple Example of a Comparison Involving Quantal Data. D. R. Cox, Biometrika, Vol. 53, No. 1/2. (Jun., 1966), pp. 215-220.

-3- Post-hoc heuristics for the Mantel-Haenszel estimator

[*Expository, with two items, near the end, left as exercises.*]

How did Mantel come up with the form for his estimator? In his paper he proposes several forms, but favours the one we know today. Some people say it was simply 'Mantel genius', but can we mere mortals find any justification for it, even post-hoc?

Breslow and Day Volume II page 109, when presenting the modification of it for situations where the amounts of exposed and unexposed Population-Time in the base are known (or, at worst, their ratio is known), refer to a 1982 paper by Clayton [available under Resources]. They say that Clayton showed that the ψ_{MH} in equation 3.16 "arises at the first stage of iteration of one of the computational methods used to find the maximum likelihood estimate." They didn't say whether it was $\psi_{ML-condn'l}$ or $\psi_{ML-uncondn'l}$, but in light of their statement a few lines further up the page, it doesn't matter, since the two are identical under an underlying Poisson model.

Mantel's version was for 'case-control' data, where, within each stratum, the observed split $d_1 : d_0$ in the denominator ('control') series is taken as an estimate of the split $PT_1 : PT_0$ in the base. For now, let's look at the easier one where that split is known without any sampling error. Moreover, let's take the *conditional* approach, where in stratum j, a total of $c_j = c_{j1} + c_{jo}$ exposed and unexposed cases arose from PT_{j1} and PT_{j0} amounts of population time respectively, and so, where

$$c_{j1} \mid c_j \ \sim \ \text{Binomial}(c_j, \pi_j = \{\psi \times PT_{j1}\}/\{\psi \times PT_{j1} + PT_{j0}\}).$$

For compactness, and looking ahead, refer to $1/\{\psi \times PT_{j1} + PT_{j0}\}$ as W_j, and drop the P from the PT.

The (binomial) likelihood contribution from stratum j is thus

$$\{\psi \times T_{j1} \times W_j\}^{c_{j1}} \times W_j^{c_{j0}} \quad \propto \ \psi^{c_{j1}} \times W_j^{c_j}.$$

So,

$$\log L_j = c_{j1} \log \psi + c_j \log W_j,$$

and so

$$d \log L_j / d\psi = \frac{c_{j1}}{\psi} + c_j \frac{d \log W_j}{d\psi} = \frac{c_{j1}}{\psi} + \frac{c_j}{W_j} \times (-1) \times W_j^2 \times T_{j1} = \frac{c_{j1}}{\psi} - c_j T_{j1} W_j.$$

Thus the estimating equation is

$$\Sigma_j c_{j1} / \hat{\psi} = \Sigma_j c_j T_{j1} W_j,$$

or, as one would expect,

$$\Sigma_j \ c_{j1} \ = \ \Sigma_j \ c_j \times \hat{\psi} \times T_{j1} \times W_j \ = \ \Sigma_j \ \hat{c}_{j1}.$$

Exercise (i): Split up c_j into $c_{j1} + c_{j0}$ and rewrite this expression as

$$\hat{\psi}_{ML} = \frac{\Sigma_j \ W_j \times c_{j1} \times T_{j0}}{\Sigma_j \ W_j \times c_{j0} \times T_{j1}}.$$

With this reformulation, we can, as Clayton 1982 explains, obtain $\hat{\psi}_{ML}$ by solving this equation by iterative refinements of the weights

$$W_j = 1/(\psi \times T_{j1} + T_{j0}), \text{ starting from } \psi = 1.$$

Clayton continues [in our notation] ...

> The first step of this iteration itself provides a fully consistent estimate, although it is only fully efficient in the neighbourhood of $\psi = 1$. This first-step estimator is closely related[12] to the Mantel-Haenszel estimator of the common odds-ratio in the $2 \times 2 \times k$ table, and the fact that the first stage of the procedure leads to quite a good estimate means that convergence is very rapid; a single refinement stage is all that will be required, except when ψ is very far from 1. The standard error of $\log \hat{\psi}$ is given by the expression
>
> $$S.E.(\log \hat{\psi}) = \{\Sigma_j \ W_j^2 \times \psi \times T_{j1} \times T_{j0} \times (c_{j1} + c_{j0})\}^{-1/2}.$$
>
> where W_j are the final weights. Of course, if the first step estimate is used, the first value of this expression gives the null s.e. of the log of the estimate, so that a test of $H_0 : \psi = 1$ may be constructed. An alternative test will be discussed immediately below.

Exercise (ii): The expression for $S.E.(\log \hat{\psi})$ looks different from that implied by equation 3.17 on page 109 of Breslow and Day Volume II. Can you reconcile them, or are they different?

[12] JH would argue that the only difference is that in a case-control study, a realization, $d_1 : d_0$, of a binomial variable is used to estimate the $PT_1 : PT_0$ ratio. Otherwise, structurally, the formulae for the known-PT and estimated -PT versions are the same.

APPENDIX: Conditional MLE of common OR (ψ) from k 2×2 tables.

In the j^{th} stratum, with fixed marginal frequencies $\underline{m} = \{c = c_{j1} + c_{j0},\ d = d_{j1}+d_{j0},\ c_{j1}+d_{j1},\ c_{j0}+d_{j0}\}$, the one independent random variable (arbitrarily c_{j1}, say), has a Non-central Hypergeometric (NCH) distribution, i.e., $c_{j1} \sim$ NCH($\underline{m}, \psi$). We will drop the j as it will be obvious from the context. This means that in a stratum,

$$pr[c_1|\underline{m}, \psi] = B[c_1] \times B[c - c_1] \times \psi^{C_1}/P[\psi],$$

where $B[c_1]$ and $B[c_1]$ are binomial coefficients, and $P[\psi]$ is the (normalizing) polynomial $\sum B[i] \times B[c-i] \times \psi^i$, and the summation is over the values of i compatible with the marginal frequencies.

Then, the log-likelihood contribution from the stratum can be written as

$$\log L = c_1 \log \psi - \log P[\psi]$$

and so, with interest in $\beta = \log \psi$ rather than ψ,

$$\frac{d \log L}{d\beta} = c_1 \frac{d \log \psi}{d\beta} - \frac{d \log P[\psi]}{d\beta} = c_1 \frac{d \log \psi}{d\psi}\frac{d\psi}{d\beta} - \frac{1}{P[\psi]}\frac{dP[\psi]}{d\psi}\frac{d\psi}{d\beta}.$$

Leaving the components in terms of ψ when it suits to keep expressions compact, and shortening $B[i] \times B[c-i]$ to B_i, we have:

$$d \log \psi / d\beta = 1; \qquad d\psi/d\beta = \exp \beta = \psi;$$

and

$$P'[\psi] = dP[\psi]/d\psi = \sum\ i \times B_i \times \psi^{i-1} = (\psi^{-1}) \times E[c_1|\psi] \times P[\psi].$$

Thus,

$$d \log L/d\beta = c_1 - (P[\psi])^{-1} \times \{\psi^{-1} \times E[c_1|\psi] \times P[\psi]\} \times \psi = c_1 - E[c_1|\psi].$$

Now, summing these scores over the strata, we have the estimating equation

$$\sum\ c_1 = \Sigma\ E[c_1|\psi] = \sum\ \widehat{c}_1.$$

The stratum-specific information I about β is $I[\beta] = -E[d^2 \log L/d\beta^2] = -d\{-E[c_1|\beta]\}/d\beta = E'[c_1|\beta]$. We can use the chain rule, or the derivative of a quotient, to obtain

$$E'[\beta] = \frac{dE}{d\psi}\frac{d\psi}{d\beta} = \frac{d\{(P[\psi])^{-1} \sum\ i \times B_i \times \psi^i\}}{d\psi}\frac{d\psi}{d\beta}$$

Using the chain rule one more time, we calculate that $E'[\beta]$ is

$$\left\{(P[\psi])^{-1} \times \{\sum i^2 \times B_i \times \psi^{i-1}\} - \{\sum\ i \times B_i \times \psi^i\}(P[\psi])^{-2}P'[\psi]\right\} \times \psi.$$

This can be simplified to

$$E'[\beta] = E[\{c_1|\psi\}^2] - \left\{\{\sum\ i \times B_i \times \psi^i\}(P[\psi])^{-1}\right\}\left\{(P[\psi])^{-1}P'[\psi]\ \psi\right\},$$

and further still, using the representation of $P'[\psi]$ above, to

$$I[\beta] = E'[\beta] = E[\{c_1|\psi\}^2] - \{E[\{c_1|\psi\}]\}^2 = Var[c_1|\psi].$$

[*there may be more elegant ways to derive the above expressions.*]

Newton-Raphson Iteration towards $\hat{\beta}_{ML-condn'l}$

Define the score function $U(\beta) = \sum_j d \log L/d\beta = \sum_j \{c_{j1} - E[c_{j1}|\beta]\}$, and the information function $I(\beta) = -\sum_j d^2 \log L/d\beta^2 = \sum_j var[c_{j1}|\beta]$.

Then, starting with $\hat{\beta}^0 = 0$, so that $\psi = 1$, we have the iterations:

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + U[\hat{\beta}^{(k)}]/I[\hat{\beta}^{(k)}].$$

As Breslow and Day hint at, each (stratum-specific) $E[c_{j1}\psi]$ and $Var[c_{j1}|\psi]$ pair involves a different *polynomial* P and so the computing challenge is to evaluate the exact expectation and variance when P involves products of large binomial coefficients, and large powers of ψ. One way to avoid unnecessarily large powers is to not always focus on the upper left cell frequency, but on the (stratum-specific) cell (possibly different in each stratum) whose possible values begin at zero.

For example, in the Estrogen and breast cancer dataset, the number of cases (c) in a riskset never exceeds 4, and the numbers of exposed subjects is never less than 36, so it makes sense to focus on the range for the number of exposed cases in the stratum.

Some `R` code that keeps the polynomials of as low a degree as possible is available under 'Software, computer code' in Resources. There was however no attempt to keep the products of binomial coefficients from causing numerical problems, and jh has not tested how large the largest marginal frequencies can be. Mind you, if one set of margins is large, and the other small, we are closer to the situation, with known PT's, discussed in 1b above.