Chapter 17
Estimating the Rate Ratio

Tabular methods

Cohort studies lend themselves to estimating the rate ratio, a measure of effect that is deficiency free or nearly so (chapter 3). To show how this key parameter can be estimated, I will use an example from a cohort of 15,712 people at baseline, 391of whom fell victims to ischemic stroke during an average follow up of 10 years. Many possible causal variables may be proposed, but in the interest of simplicity we will consider only two: hypertension status (playing the role of an exposure) and age (a confounder or an effect modifier). Hypertension status usually belongs to the world of binary variables, whereas age was categorized into four groups for didactic reasons. Table 17–1 shows relevant data. Examine this table carefully; it is the foundation of what follows.

Table 17–1. Number of participants, person-years at risk, stroke cases, rates and rate ratios, by hypertension status and age group.

	Hypertension					Normotension				Rate Ratio
Age Group	Number of People	Person-years at risk	Number of Strokes	Rate (per 10,000)		Number of People	Person-years at risk	Number of Strokes	Rate (per 10,000)	
45-49	1,046	10,329	39	37.8		3,173	32,144	13	4.0	9.3
50-54	1,299	12,669	45	35.5		2,768	28,022	24	8.6	4.1
55-59	1,476	14,053	75	53.4		2,364	23,411	44	18.8	2.8
60-64	1,683	15,243	108	70.9		1,903	18,409	43	23.4	3.0
All	5,504	52,294	267	51.1		10,208	101,986	124	12.2	4.2

About one third of the participants (5,504) were classified as having hypertension. This part of the cohort has "contributed" 52,294 person-years at risk and, unfortunately, 267 strokes. The remainder of the cohort (10,208 participants with normal blood pressure) has accounted for 101,986 person-years at risk and 124 strokes. You will find these numbers in the last row of Table 17–1, and again in Table 17–2.

Table 17–2. Number of strokes and person-years at risk, by hypertension status

	Number of strokes	Person-years at risk
Hypertension	267 (a)	52,294 (N_1)
Normotension	124 (b)	101,986 (N_2)
All	391	154,280

The marginal (crude) association is described by the rate ratio:
Rate Ratio = $(a/N_1)/(b/N_2)$ = (267/52,294)/(124/101,986) = 4.2

Neither you nor I would be willing to assume that the marginal association estimates the hypertension effect on stroke, because we can think of several confounding paths: ageinduced, for example. For this reason, both the estimate and the standard error of the estimator behind it should be declared meaningless from a causal perspective. Nonetheless, I will compute the standard error to illustrate the method you would use if no conditioning on confounders were needed—say, if the causal assignments were determined at random.

The standard error around the log of the rate ratio is a function of the number of events in each group (here, the number of strokes). Following the notation of Table 17-2, it may be estimated as follows.

$$\text{SE}[\log(\text{rate ratio})] = \sqrt{(1/a)+(1/b)} = \sqrt{(1/267)+(1/124)} = 0.1087$$

Using the standard error, we can compute three kinds of 95% confidence limits:

CI for the log(rate ratio): log(4.2)$\pm$1.96 x 0.1087 = 1.435$\pm$1.96 x 0.1087 = [1.222, 1.648]

CI for the rate ratio: [exp(1.222), exp(1.648)] = [3.4, 5.2]

Confidence limit ratio (CLR) for the rate ratio: 5.2/3.4 = 1.5

Let's consider next the role that age might play in our attempt to estimate the hypertension effect. Looking at Table 17–1 again, we see that the age-specific rate ratios range from 2.8 to 9.3, so effect modification by age cannot be dismissed on this scale (and perhaps on the additive scale, too). Nonetheless, I will assume homogeneity of the underlying causal parameter from which these estimates arose, and treat age as a confounder in line with a naïve causal diagram (Figure 17–1).

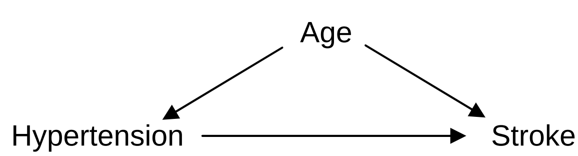

Figure 17–1. A causal diagram showing confounding by age

To estimate the effect of hypertension on stroke, you should condition on age. For example, stratify the sample on age group and calculate a weighted average of four age-specific rate ratios. If we use the subscript "*i*" to denote the *i*-th stratum, and denote the weight by "w", then

$$\text{RR(adjusted)} = \frac{\sum w_i RR_i}{\sum w_i}$$

To show the computation, I extracted relevant data from the rows of Table 17–1 and created four tables, one per age group (Table 17–3.)

Table 17–3(a–d). Age-specific relation of hypertension status and stroke

a. Age group: 45-49 years

	Number of strokes	Person-years at risk	Rate Ratio	Weight
Hypertension	39	10,329	Ref.	
Normotension	13	32,144	9.3	3.16

b. Age group: 50-54 years

	Number of strokes	Person-years at risk	Rate Ratio	Weight
Hypertension	45	12,669	Ref.	
Normotension	24	28,022	4.1	7.47

c. Age group: 55-59 years

	Number of strokes	Person-years at risk	Rate Ratio	Weight
Hypertension	75	14,053	Ref.	
Normotension	44	23,411	2.8	16.50

d. Age group: 60-64 years

	Number of strokes	Person-years at risk	Rate Ratio	Weight
Hypertension	108	15,243	Ref.	
Normotension	43	18,409	3.0	19.48

How did I compute the weights?

We naturally expect the weight of the stratum-specific rate ratio to be inversely related to the variance of the stratum-specific estimator: the larger the variance, the smaller should be the weight. Mathematical details aside, the following formula, which was proposed by Mantel and Haenszel, approximates that kind of weight for each age group.

$$\frac{\text{Number of strokes among } \textbf{normotensives} \times \text{Person-years at risk of } \textbf{hypertensives}}{\textbf{Total person-years} \text{ at risk in that age group}}$$

For instance, the weight for the oldest group is

(43 x 15,243) / (15,243+18,409) = 19.48

and it is much larger than the corresponding weight for the youngest group (3.16). Such a ranking appeals to our intuition. The oldest group has contributed more "data", more strokes have occurred in that group, so its rate ratio (3.0) should have greater influence on the weighted average than, for example, the rate ratio of the youngest group (9.3). Again, we are assuming that both numbers (in fact, all four rate ratios) estimate a single, common causal parameter and that no other confounders exist.

Applying the generic formula for a weighted average, we can calculate the conditional rate ratio according to the Mantel-Haenszel formula ($RR_{M\text{-}H}$).

$$RR_{M\text{-}H}(\text{age-adjusted}) = \frac{\sum w_i RR_i}{\sum w_i} = \frac{3.16x9.3 + 7.47x4.1 + 16.50x2.8 + 19.48x3.0}{3.16 + 7.47 + 16.50 + 19.48} = 3.6$$

Recall that the marginal rate ratio was 4.2. Conditioning on age has, therefore, attenuated the association between hypertension and stroke—as expected. Since hypertensives were older than normotensives, part of the marginal association has embedded the age effect on hypertension and stroke.

With a little notation and simple algebra, it is possible to express the Mantel-Haenszel formula differently. First, we display the data for the *i*-th stratum of the confounder by adding the subscript "*i*" (Table 17–4).

Table 17–4. Number of events and person-time at risk in the *i*-th stratum

	Number of events	Person-time at risk
Exposed	a_i	N_{1i}
Unexposed	b_i	N_{2i}
All		N_{Ti}

As before, the weight of the *i*-th stratum is given by: $(b_i \, x \, N_{1i}) / N_{Ti}$

The weighted average is then,

$$RR_{M\text{-}H} = \frac{\sum w_i RR_i}{\sum w_i} = \frac{\sum (b_i N_{1i} / N_{Ti}) x (a_i / N_{1i}) / (b_i / N_{2i})}{\sum (b_i N_{1i} / N_{Ti})} = \frac{\sum a_i N_{2i} / N_{Ti}}{\sum b_i N_{1i} / N_{Ti}}$$

Applying the formula on the right hand side to our example, we get the same adjusted rate ratio:

$$RR_{M\text{-}H} = \frac{(39x32{,}144/42{,}473) + (45x28{,}022/40{,}691) + (75x23{,}411/37{,}464) + (108x18{,}409/33{,}652)}{(13x10{,}329/42{,}473) + (24x12{,}669/40{,}691) + (44x14{,}053/37{,}464) + (43x15{,}243/33{,}652)} = 3.6$$

In this version of the Mantel-Haenszel formula, we circumvent the need to compute stratum-specific rate ratios and stratum-specific weights. Although the calculation is simpler and faster than the original math, you are paying a double price for the shortcut: first, you don't get to see the rate ratios that make up the average. Second, you don't get to see the relative weights.

If you wish to compute the standard error of the log ($RR_{M\text{-}H}$), take the square root of the following expression:

$$\text{Var}[\log(\text{RR}_{\text{M-H}})] = \frac{\sum (a_i + b_i) N_{1i} N_{2i} / T_i^2}{(\sum a_i N_{2i} / T_i)(\sum b_i N_{1i} / T_i)} \text{ [= 0.011291]}$$

SE $[\log(\text{RR}_{\text{M-H}})] = \sqrt{0.011291} = 0.1063$

Or perhaps it is time to switch to Poisson regression and read these numbers off a printout…

Poisson regression

Poisson regression is one of two regression models by which we can estimate marginal rate ratios and conditional (adjusted) rate ratios. (The other is Cox regression.) I will first develop the theory behind the model and then illustrate the SAS code using, again, the example of hypertension and stroke.

Let E be a binary exposure: 1=EXPOSED; 0=UNEXPOSED. Other covariates and interaction terms may be added, but are avoided to simplify notation. In our example of stroke, the exposure is hypertension status, and the goal is to estimate the rate ratio for the contrast between exposed (hypertensives) and unexposed (normotensives).

Let the Greek letter λ stand for "rate". On first try, we might specify the following regression model:

$$\lambda = \beta_0 + \beta_1 \text{E}$$

The model is reasonable but the coefficient of E estimates the rate *difference*, not the rate *ratio*. If we wish to estimate the rate ratio, we should substitute $\log(\lambda)$ for λ.

$$\log(\lambda) = \beta_0 + \beta_1 \text{E} \qquad \text{(Equation 17–1)}$$

In equation 17–1, the coefficient of the exposure is the log of the rate ratio, analogous to the log of the odds ratio in logistic regression. Therefore, Rate Ratio = $\exp(\beta_1)$

Since "rate" is defined as the number of events (which I will call "μ") per person-time at risk (which I will call "N"), we may write "$\lambda = \mu / \text{N}$", and rewrite equation 17–1 as follows:

$$\log(\mu / \text{N}) = \beta_0 + \beta_1 \text{E}$$

A little more algebra takes us to the following equations:

$$\log(\mu) - \log(\text{N}) = \beta_0 + \beta_1 \text{E}$$

$$\log(\mu) = \beta_0 + \beta_1 \text{E} + \log(\text{N}) \qquad \text{(Equation 17–2)}$$

$$\mu = \exp[\beta_0 + \beta_1 \text{E} + \log(\text{N})] \qquad \text{(Equation 17–3)}$$

Notice that the coefficients in equation 17–2 or equation 17–3 are identical to the coefficients in equation 17–1. Therefore, if we find a way to estimate the parameters of the last two equations, the rate ratio will be in our hands: $\exp(\beta_1)$.

To estimate β_0 and β_1 in equation 17–3 (or 17–2), we will have to construct a likelihood function (L), called the Poisson likelihood, analogous to the binomial likelihood, which we used to estimate the coefficients of a logistic regression model. Once we succeed in expressing L as a function of β_0 and β_1, we will search for the maximum likelihood estimates—for those values of β_0 and β_1 that generate the largest possible value of L. The road from here to the last step is a little long—about 4 pages—but I think it's worth following.

As always, the likelihood is defined as the probability of observing "the data". In our example of hypertension and stroke, "the data" mean 267 strokes during 52,294 person-years at risk of hypertensives and 124 strokes during 101,986 person-years at risk of normotensives. Since the occurrence of stroke in one group is independent of its occurrence in the other, the probability of observing both counts—the likelihood—is the product of two independent probabilities.

$L = \Pr(Y=267) \times \Pr(Y=124)$

What, then, are these probabilities? What formula may we use to compute them?

That's the place where an interesting probability distribution enters the story.

Poisson probability distribution

A few hundred years ago, Simeon Poisson proposed that the probability of observing "r" events might follow a "strange-looking" formula:

$$\Pr(Y=r) = e^{-\mu} (\mu)^{r} / r! \qquad \text{(Equation 17–4)}$$

For example, the probability of observing 267 strokes in our sample of hypertensives is

$$\Pr(Y=267) = e^{-\mu} (\mu)^{267} / 267!$$

Let's examine slowly the content of the right hand side of these equations: "e" is that well known irrational number (2.718…); "r" is the number of events we specify, such as 267; and r! (r factorial) is short for multiplication of sequential integers (1x2x3x…r). But what is μ in this equation? Well, μ is the number of events (here, the number of strokes) we *expect* to observe in our sample—the most probable number of events we expect to observe. To use an example from the gambling world: Probability calculations could lead us to expect two winners of the lottery (μ=2) among one million lottery buyers, but we might observe one winner (r=1) or fifty winners (r=50), each with a certain probability.

What, then, determines the value of μ, the number of events we expect to observe?

The answer should become apparent after recalling the formula for a rate "$\lambda = \mu / N$", and rewriting it as "$\mu = \lambda \times N$". Both the person-time at risk (N) and the rate (λ) determine the expected number of events (μ). The larger is the person-time at risk and the larger the rate, the more events are expected to occur. As we know, the person-time at-risk is largely determined by our study design, namely, the available follow up time, but what factors set the value of λ, the rate?

That question was addressed in chapter 3. In an indeterministic world the rate reflects the strength of all causal forces behind the event in question, which push toward realization of the effect. We do not know, of course, how these forces determine the rate, or even the name of every causal variable, but our naïve regression model (equation 17–1) has assumed a simple mathematical relation between a single cause (E) and the log of the rate (λ): $\log(\lambda) = \beta_0 + \beta_1 E$

It is not crucial for you to understand the shape of the Poisson distribution, but it might be interesting. Let's compute several Poisson probabilities for λ=0.003 and N=1,000 person-years at risk. On these assumptions, the expected number of events (μ) is 3 ($\mu = \lambda \times N = 0.003 \times 1{,}000 = 3$). Enter μ=3 into equation 17–4 and you get the formula for the probability of observing any number of events (r) you would like to specify.

$$\Pr(Y=r) = e^{-3}\,(3)^{r} / r!$$

For r= 0, 1, 2, 3,…,10 , we get the following Poisson probabilities:

$\Pr(Y=0) = e^{-3}\,(3)^{0} / 0! = 0.05$ (0!=1 by definition)

$\Pr(Y=1) = e^{-3}\,(3)^{1} / 1! = 0.15$

$\Pr(Y=2) = e^{-3}\,(3)^{2} / 2! = 0.22$

$\Pr(Y=3) = e^{-3}\,(3)^{3} / 3! = 0.22$

$\Pr(Y=4) = e^{-3}\,(3)^{4} / 4! = 0.17$

$\Pr(Y=5) = e^{-3}\,(3)^{5} / 5! = 0.10$

$\Pr(Y=6) = e^{-3}\,(3)^{6} / 6! = 0.05$

$\Pr(Y=7) = e^{-3}\,(3)^{7} / 7! = 0.02$

$\Pr(Y=8) = e^{-3}\,(3)^{8} / 8! = 0.008$

$\Pr(Y=9) = e^{-3}\,(3)^{9} / 9! = 0.003$

$\Pr(Y=10) = e^{-3}\,(3)^{10} / 10! = 0.001$

For example, with an underlying rate of 0.003, the probability of observing 1 event in a cohort of 1,000 person-years at risk is 0.15, whereas the probability of observing 10 events is only 0.001, a very small chance. Figure 17–1 displays the above probabilities.

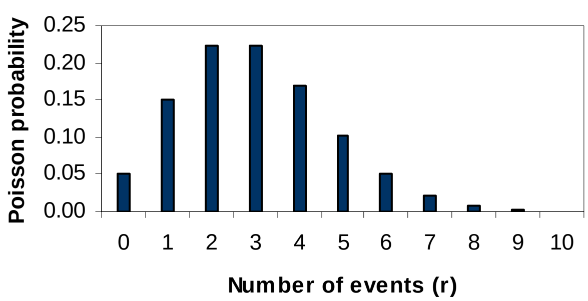

Figure 17–1. The Poisson probability distribution for $\mu=3$

Since the Poisson distribution is a probability density function (chapter 8), the height of the bars sum to 1. In other words, Pr $(r \geq 0)=1$ because we are certain to observe "something" (either no event or some number of events.) But as you can see, the distribution is skewed to the right, having a long thin tail. There is nothing surprising here. When the event does not happen often (small λ, low rate) and the person-time at risk is modest, it is improbable to observe many events. Notice also that the maximum probability is reached when the number of events is 2 or 3—at or near the expected value ($\mu=3$).

Poisson likelihood function

After a detour through the Poisson probability distribution, let's return to the example of hypertension and stroke, and write the Poisson likelihood for the observed data, which was our original goal. Again, the likelihood in this case is the product of two independent probabilities, each of which is assumed to be a Poisson probability $(e^{-\mu} (\mu)^r / r!)$

$$L = \Pr(Y=267) \quad \times \quad \Pr(Y=124)$$

$$= e^{-\mu_1} (\mu_1)^{267} / 267! \quad \times \quad e^{-\mu_2} (\mu_2)^{124} / 124!$$

Keep in mind the meaning of μ_1 and μ_2 in the expression above: They are the *expected* number of strokes in hypertensives and normotensives, respectively. And they should be different for two reasons: the person-time at risk is different (52,294 person-years of hypertensive people versus 101,986 person-years of normotensive people) and the underlying rate may be different because of causal variables, such as hypertension status.

We have already seen in logistic regression that it is easier to work with the log-likelihood function than with the likelihood itself. So we'll take the log of the last expression:

$$\text{Log–L} = -\mu_1 + 267 \log(\mu_1) - \log(267!) + [-\mu_2 + 124 \log(\mu_2) - \log(124!)]$$

$$= 267 \log(\mu_1) - \mu_1 + 124 \log(\mu_2) - \mu_2 - \log(267!) - \log(124!)$$

We have already seen in the context of logistic regression that constant terms, such as log(267!) and log(124!), do not affect the computation of maximum likelihood estimates. Everyone omits them to simplify mathematical expressions and so will we:

$$\text{Log–L (revised)} = 267 \log(\mu_1) - \mu_1 + 124 \log(\mu_2) - \mu_2 \qquad \text{(Equation 17–5)}$$

So far we expressed the likelihood as a function of μ_1 and μ_2. Now it's time to invoke the Poisson regression model itself and to specify μ_1 and μ_2—the expected number of strokes in each group—as a function of the exposure variable (hypertension status) and the person-years at risk. Specifically, recall that our regression model has assumed the following mathematical relation of E and N with μ:

$$\log(\mu) = \beta_0 + \beta_1 \, \mathrm{E} + \log(\mathrm{N}) \qquad \text{(Equation 17–2)}$$

$$\mu = \exp[\beta_0 + \beta_1 \, \mathrm{E} + \log(\mathrm{N})] \qquad \text{(Equation 17–3)}$$

For the group of hypertensives, E=1 and N=52,294, so $\log(\mu_1)$ and μ_1 are as follows:

$$\log(\mu_1) = \beta_0 + \beta_1 + \log(52294)$$

$$\mu_1 = \exp[\beta_0 + \beta_1 + \log(52294)]$$

For the group of normotensives, E=0 and N=101,986, so $\log(\mu_2)$ and μ_2 are as follows:

$$\log(\mu_2) = \beta_0 + \log(101986)$$

$$\mu_2 = \exp[\beta_0 + \log(101986)]$$

Plugging these expressions of $\log(\mu_1)$, μ_1, $\log(\mu_2)$, and μ_2 into the Poisson log-likelihood function (equation 17–5), we get the following:

$$\begin{aligned}
\text{Log–L (revised)} &= 267 \log(\mu_1) - \mu_1 + 124 \log(\mu_2) - \mu_2 \\
&= 267\,[\beta_0 + \beta_1 + \log(52294)] - \exp[\beta_0 + \beta_1 + \log(52294)] \\
&\qquad + 124\,[\beta_0 + \log(101986)] - \exp[\beta_0 + \log(101986)] \\
&= 267\,(\beta_0 + \beta_1) + 267 \log(52294) - \exp(\beta_0 + \beta_1)\mathrm{x}52294 \\
&\qquad + 124\,\beta_0 + 124 \log(101986) - \exp(\beta_0)\mathrm{x}101986
\end{aligned}$$

Again, the addition or subtraction of constants, such as "267 log((52294)", does not affect the maximum likelihood estimates; it just shift the entire function up or down. To simplify, we'll omit all constants and combine some terms to get the simplest possible expression:

$$\text{Log–L (revised)} = (267{+}124)\beta_0 + 267\,\beta_1 - [\exp(\beta_0 + \beta_1)\mathrm{x}52294 + \exp(\beta_0)\mathrm{x}101986]$$

We are done! Examine the right hand side of the last equation and you will see that we finally expressed the log-likelihood as a function of β_0 and β_1, which was our goal at the beginning of this long journey. In mathematical notation: Log–L = $f(\beta_0, \beta_1)$. All that is left to do is to find the values of β_0 and β_1 that maximize the value of the function, and that can be done by iteration ("trial and error") with the help of an algorithm. Retrace the steps back to equation 17–1 and you will realize that "$\exp(\beta_1)$" estimates the rate ratio for stroke for the contrast between hypertensives and normotensives.

Let's see how SAS does it in a procedure called `PROC GENMOD`.

SAS PROC GENMOD

SAS code (first, formatting and data steps)

```
PROC FORMAT;
 VALUE htnfmt 0='Normotensive'
              1='Hypertensive';
 VALUE agefmt 1='D. 45-49'
              2='C. 50-54'
              3='B. 55-59'
              4='A. 60-64';
 run;

DATA Poisson;
 INPUT htn agegroup people personyears events;
 logPYEARS=log(personyears);

DATALINES;

0 1 3173 32144   13
0 2 2768 28022   24
0 3 2364 23411   44
0 4 1903 18409   43
1 1 1046 10329   39
1 2 1299 12669   45
1 3 1476 14053   75
1 4 1683 15243  108
;
run;
```

Rather than reading the data (Table 17–1) from a data file, I entered the numbers directly in a data step (`DATALINES`). Hypertension status and age group were coded as follows:

HTN: 1=HYPERTENSION; 0=NORMOTENSION
AGEGROUP: 1=45-49; 2=50-54; 3=55-59; 4=60-64

For a reason that will become clear shortly, I also had to create a new variable, logPYEARS, which is the log of the person-years at risk. Next is `PROC GENMOD`.

```
PROC GENMOD;
 CLASS htn;
 MODEL events = htn / DIST=POISSON
                      LINK=LOG
                      OFFSET=logPYEARS;
ESTIMATE 'Beta htn' htn 1 -1/ exp;
```

```
FORMAT HTN htnfmt.;
run;
```

To follow the logic of the `PROC GENMOD` code, recall the Poisson regression equation (equation 17–2):

$$\log(\mu) = \beta_0 + \beta_1 \text{E} + \log(\text{N}) \quad \text{and compare it to the "model statement"}$$

```
MODEL events = htn / DIST=POISSON
                     LINK=LOG
                     OFFSET=logPYEARS;
```

On the left hand side of the "model statement", you find the variable EVENTS, the number of strokes, which is assumed to follow a Poisson probability distribution (`DISTRIBUTION=POISSON`). But as the regression model shows, we should request the software to predict the log of that number (`LINK=LOG`).

What is the purpose of the code "`OFFSET=logPYEARS`"?

Notice that log(N) appears as a regressor on the right hand side of the Poisson regression equation, so we should have included it somehow in the "model statement". If we wrote, however, "`MODEL events = htn logPYEARS`", SAS would have estimated a coefficient for this variable, too. We don't want that to happen—we did <u>not</u> specify the equation as "$\log(\mu) = \beta_0 + \beta_1 \text{E} + \beta_3 \log(\text{N})$" but as "$\log(\mu) = \beta_0 + \beta_1 \text{E} + \log(\text{N})$". The "coefficient" of log(N) should be 1.

The option "`OFFSET=logPYEARS`" serves that purpose. `OFFSET` means that logPYEARS is a special regression variable whose coefficient should not be estimated; it must be 1 (β_3=1).

Selected SAS output

```
                 The GENMOD Procedure
                  Model Information

          Data Set                WORK.POISSON
          Distribution                 Poisson
          Link Function                    Log
          Dependent Variable            events
          Offset Variable            logPYEARS
          Observations Used                  8

                Class Level Information

    Class      Levels    Values

    htn             2    Hypertensive Normotensive

                  Algorithm converged.
```

```
                         Analysis Of Parameter Estimates

                                             Standard       Wald 95%
Parameter                 DF   Estimate        Error   Confidence Limits

Intercept                  1    -6.7123       0.0898   -6.8883   -6.5363
htn         Hypertensive   1     1.4349       0.1087    1.2219    1.6479
htn         Normotensive   0     0.0000       0.0000    0.0000    0.0000

                             Contrast Estimate Results

                                Standard
 Label            Estimate        Error              Confidence Limits

 Beta htn           1.4349       0.1087              1.2219     1.6479
 Exp(Beta htn)      4.1993                           3.3937     5.1961
```

Regression equation: log (stroke rate) = –6.7123 + 1.4349 HTN

The output I selected is self-explanatory. After exponentiating the coefficient of the hypertension variable, we get a rate ratio of 4.2. Compare this rate ratio and its standard error to the numbers we computed by hand at the very beginning of this chapter: The two methods have produced identical results. Why do two vastly different mathematical trails lead to identical results? Why should the most likely estimate from a Poisson likelihood function, which is founded on a strange-looking Poisson probability, precisely match the simple rate ratio we have quickly computed by hand? I don't know the answer, but it's a good opportunity to ponder again about the mathematical fabric of the universe. Have we invented statistics to discover causal connections or have we discovered the statistics with which Nature invented causal connections?

Next, we will condition the association of hypertension and stroke on age by adding the variable AGEGROUP to the "model statement". Recall that in tabular methods, we conditioned on age by stratification, followed by the computation of a weighted average (RR_{M-H}) of the age-specific rate ratios. In regression, conditioning is done in a black box; we get to see only the final result.

There is more than one way to model the 4-level age variable. If we add AGEGROUP to the "class statement", SAS will replace that variable with three "dummy variables", selecting one age group as the reference (Table 17–5).

Table 17–5. Substituting 3 "dummy variables" for AGEGROUP

AGEGROUP	AGE50-54	AGE55-59	AGE60-64
1 (45-49 years)	0	0	0
2 (50-54 years)	1	0	0
3 (55-59 years)	0	1	0
4 (60-64 years)	0	0	1

```
PROC GENMOD;
 CLASS htn agegroup;
 MODEL events = htn agegroup / DIST=POISSON
                               LINK=LOG
                               OFFSET=logPYEARS;
ESTIMATE 'Beta htn' htn 1 -1/ exp;
FORMAT htn htnfmt.;
FORMAT agegroup agefmt.;

run;
```

Selected SAS output

```
                            The GENMOD Procedure

                             Model Information

                    Data Set                WORK.POISSON
                    Distribution                 Poisson
                    Link Function                    Log
                    Dependent Variable            events
                    Offset Variable            logPYEARS
                    Observations Used                  8


                          Class Level Information

         Class          Levels    Values

         htn                 2    Hypertensive Normotensive
         agegroup            4    A. 60-64 B. 55-59 C. 50-54 D. 45-49


  Algorithm converged.


                       Analysis Of Parameter Estimates

                                             Standard       Wald 95%
 Parameter                  DF   Estimate       Error   Confidence Limits

 Intercept                   1    -7.2080      0.1510   -7.5039   -6.9122
 htn        Hypertensive     1     1.3044      0.1100    1.0888    1.5200
 htn        Normotensive     0     0.0000      0.0000    0.0000    0.0000
 agegroup   A. 60-64         1     1.0056      0.1625    0.6872    1.3240
```

agegroup	B. 55-59	1	0.7592	0.1670	0.4319	1.0865
agegroup	C. 50-54	1	0.2207	0.1839	-0.1397	0.5811
agegroup	D. 45-49	0	0.0000	0.0000	0.0000	0.0000

Contrast Estimate Results

Label	Estimate	Standard Error	Confidence Limits	
Beta htn	1.3044	0.1100	1.0888	1.5200
Exp(Beta htn)	3.6855		2.9708	4.5721

log (stroke rate) = –7.208 + 1.3044 HTN +
1.0056 AGE60-64 + 0.7592 AGE55-59 + 0.2207 AGE50-54

Just like conditioning in tabular methods, adding age to the regression model has attenuated the association between hypertension and stroke. The age-adjusted rate ratio from this model, exp(1.3044)=3.7, is similar to the Mantel-Haenszel rate ratio (3.6). Assuming that no other conditioning is needed, you may report the 95% CI (2.9 to 4.6) and the 95% CLR (4.57/2.97=1.5). (In scientific inquiry, I would not. The estimator and its standard error are still useless from a causal perspective. It is easy to propose other confounding paths.)

Poisson regression and person-based data file

Although we developed the Poisson likelihood for group data (Table 17–1), the content of that table was obtained by observing individuals. Each person has contributed years at risk (a value of the variable N) and event status over follow up: Y=1, if suffered a stroke or Y=0, if remained stroke-free. These data were then summarized for hypertensives and normotensives and for four age groups. What would the likelihood function look like if we were to use the original, person-based, data?

In that case the likelihood function is not necessarily "Pr (Y=267) x Pr (Y=124)", but may be expressed as the product of 15,712 individual probabilities—the size of our cohort. Depending on the fate of each cohort member, he or she "contributes" a probability of having suffered a stroke (Pr (Y=1)), or of having remained stroke-free (Pr (Y=0)). The likelihood, L, is therefore

$L = Pr_1 \times Pr_2 \times Pr_3 \times \ldots \times P_{15712}$

Again, let's switch to the log-likelihood function because it is simpler to work on that scale. As you know, the log of the product of terms is equal to the sum of the log of each term.

Log-L = log $(Pr_1 \times Pr_2 \times Pr_3 \times \ldots x P_{15712})$ = log (Pr_1) + log (Pr_2) + log (Pr_3) +…+ log (Pr_{15712})

Next, we'll assume that each of these 15,712 probabilities is a Poisson probability (more on that assumption later):

$$\Pr(Y=r) = e^{-\mu} (\mu)^{r} / r!$$

For a single person, however, "r" can take only two values: r=1, if the person had suffered a stroke or r=0, if the person had not.

If the person had suffered a stroke, r=1 : $\Pr(Y=1) = e^{-\mu}\mu^1/1! = e^{-\mu}\mu$

And the log of that probability is $\log[\Pr(Y=1)] = \boldsymbol{\log(\mu) - \mu}$

If the person remained stroke-free, r=0 : $\Pr(Y=0) = e^{-\mu}\mu^0/0! = e^{-\mu}$

And the log of that probability is $\log[\Pr(Y=0)] = \boldsymbol{-\mu}$

Therefore, the log likelihood is the sum of two kinds of log of probability:

Every stroke victim contributes to the summation "$\log(\mu) - \mu$", and there are 391 such people. Similarly, those who escaped that fate contribute "$-\mu$", and there are 15,321 such people. In semi-formal notation:

$$\text{Log-L} = \sum^{391}[\log(\mu) - \mu] + \sum^{15,321}[-\mu] \qquad \text{Equation (17–6)}$$

All that is left to do is to replace μ in the last equation with expressions that contain β_0 and β_1, namely, with the right hand side of the Poisson regression equations:

$$\log(\mu) = \beta_0 + \beta_1 E + \log(N) \qquad \text{(Equation 17–3)}$$

$$\mu = \exp[\beta_0 + \beta_1 E + \log(N)] \qquad \text{(Equation 17–4)}$$

Here are two examples that illustrate the replacement:

- If Mr. Smith was hypertensive (E=1) and remained stroke-free during 7 years at risk (N=7), his value of $\mu = \exp[\beta_0 + \beta_1 + \log(7)]$. Since Mr. Smith is one of 15,321 people who did not suffer a stroke, his contribution to equation 17–6 would be $-\mu$, which is "$-\exp[\beta_0 + \beta_1 + \log(7)]$".
- If Ms. Jones was normotensive (E=0) and suffered a stroke after 10 years (N=10), her value of $\mu = \exp[\beta_0 + \log(10)]$, and her contribution to equation 17–6 would be $\log(\mu) - \mu$, which is "$\beta_0 + \log(10) - \exp[\beta_0 + \log(10)]$"

After summing all 15,712 replacing terms, the log-likelihood will be, again, a function of β_0 and β_1. I could have ended the story by showing a formal messy expression of the function, but it is not essential. The principles should suffice.

Since "r" is constrained to be "1" or "0", is it legitimate to fit a Poisson regression model to person-based data? After all, it is difficult to conceive a complete Poisson distribution for a single person, such as Mr. Smith: one can suffer no more than one *incident* stroke, and that is devastating enough.

The answer should be "yes" for several reasons: First, a few lines of algebra can show that the Poisson distribution is the limit of the binomial distribution when the probability of the event tends to zero and the person-time at risk tends to infinity. If we apply the Poisson distribution to a large cohort (many years at risk per person) and a rare event (low rate), we are effectively approximating a binomial probability distribution. No one would complain about using the latter for a binary dependent variable.

Second, think for a moment about a data file that contains "grouped data"—Table 17–1 for instance—and reconstruct it in your mind as a file that contains 15,712 individual records. If you are willing to apply Poisson regression to the group file, should you not be willing to do so to its person-based counterpart? Is a probability model tied to the method by which we organize the entries in a data file, or does it try to describe underlying causal reality?

Finally (and a little more abstract): although Mr. Smith (for example) has contributed one row of data (E=1, N=7, Y=0), we may view his contribution to the *right* hand side of the regression "$\mu = \exp[\beta_0 + \beta_1 E + \log(N)]$" as just one sample of many similar observations—of many Smith-like replications of E=1 and N=7. Because a theoretical collective of [E=1; N=7] can generate more than a single stroke ($\mu > 1$), we may "legitimately" invoke the Poisson probability distribution. In that abstract framework, which resonates with indeterministic causation (chapter 1), "r" is not constrained to be "0" or "1" even though it is empirically impossible to observe anything greater than r=1 in any given person.

The SAS code below was fit to the original stroke data, namely, to a data file that contained 15,712 observations. Notice two key changes: (1) The dependent variable is not EVENTS but STROKE, a binary variable, which takes the value of 1 or 0. (2) Instead of logPYEARS, I used a variable called logPY—the log of years at risk for <u>each</u> member of the cohort. The first model that I fit estimates the marginal rate ratio; the second, the so-called age-adjusted rate ratio. In the second model SAS, again, has replaced the variable AGEGROUP with three dummy variables, choosing the youngest group as the reference (Table 17–5.)

SAS code

```
PROC GENMOD;
 CLASS htn;
 MODEL stroke = htn  / DIST=POISSON
                       LINK=LOG
                       OFFSET=logPY;
ESTIMATE 'Beta htn' htn 1 -1/ exp;
run;

PROC GENMOD;
 CLASS htn agegroup;
 MODEL stroke = htn agegroup / DIST=POISSON
```

```
                                   LINK=LOG
                                   OFFSET=logPY;
ESTIMATE 'Beta htn' htn 1 -1/ exp;

run;
```

Selected SAS output

```
                              The GENMOD Procedure

                               Model Information

          Data Set                WORK.ONE
          Distribution             Poisson
          Link Function                Log
          Dependent Variable        stroke
          Offset Variable            logPY
          Observations Used          15712


                            Class Level Information

                Class      Levels    Values

                htn             2    Hypertensive Normotensive

  Algorithm converged.


                         Analysis Of Parameter Estimates

                                              Standard       Wald 95%
 Parameter                 DF   Estimate       Error   Confidence Limits

 Intercept                  1    -6.7123      0.0898   -6.8883   -6.5363
 htn        Hypertensive    1     1.4349      0.1087    1.2219    1.6479
 htn        Normotensive    0     0.0000      0.0000    0.0000    0.0000


                             Contrast Estimate Results

                             Standard
  Label            Estimate      Error              Confidence Limits

  Beta htn           1.4349     0.1087              1.2219     1.6479
  Exp(Beta htn)      4.1993                         3.3937     5.1961
```

log (stroke rate) = –6.7123 + 1.4349 HTN

The GENMOD Procedure

Model Information

Data Set	WORK.ONE
Distribution	Poisson
Link Function	Log
Dependent Variable	stroke
Offset Variable	logPY
Observations Used	15712

Class Level Information

Class	Levels	Values
htn	2	Hypertensive Normotensive
agegroup	4	A. 60-64 B. 55-59 C. 50-54 D. 45-49

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits	
Intercept		1	-7.2080	0.1510	-7.5039	-6.9121
htn	Hypertensive	1	1.3044	0.1100	1.0888	1.5200
htn	Normotensive	0	0.0000	0.0000	0.0000	0.0000
agegroup	A. 60-64	1	1.0055	0.1625	0.6871	1.3240
agegroup	B. 55-59	1	0.7592	0.1670	0.4318	1.0865
agegroup	C. 50-54	1	0.2207	0.1839	-0.1397	0.5811
agegroup	D. 45-49	0	0.0000	0.0000	0.0000	0.0000

Contrast Estimate Results

Label	Estimate	Standard Error	Confidence Limits	
Beta htn	1.3044	0.1100	1.0888	1.5200
Exp(Beta htn)	3.6854		2.9708	4.5720

$$\log(\text{stroke rate}) = -7.208 + 1.3044\ \text{HTN} + 1.0055\ \text{AGE60-64} + 0.7592\ \text{AGE55-59} + 0.2207\ \text{AGE50-54}$$

Table 17–6 compares the estimates we obtained for the hypertension "effect" by fitting Poisson regression to person-based data to those we had obtained before by tabular methods and by fitting Poisson regression to group data. The similarity is remarkable.

Table 17–6. Point estimates of the rate ratio and standard errors, by three methods of estimation

	Tabular Methods (before rounding)	Poisson Regression (group data)	Poisson Regression (person-based)
Marginal rate ratio	4.1993	4.1993	4.1993
Conditional rate ratio*	3.5707	3.6855	3.6854
Standard error**	0.1063	0.1100	0.1100

*"Age-adjusted"
**SE of log (conditional rate ratio)

Estimating the modified rate ratio

Suppose we decide that age plays the role of an effect modifier and, therefore, prefer to present age-specific rate ratios of the hypertension effect (Table 17–1). To obtain these estimates by regression, we may fit a model that contains interaction terms between age and hypertension.

In one commonly used method, the 4-level AGEGROUP variable is first replaced by three "dummy variables", choosing one age group as the reference (see Table 17–5 again). Then, we fit a model that contains three interaction terms (in addition, of course, to the "main effects"): HTN x AGE50-54, HTN x AGE55-59, and HTN x AGE60-64. Here is that model:

$$\begin{aligned}\log(\text{stroke rate}) = \beta_0 &+ \beta_1\ \text{HTN} \\ &+ \beta_2\ \text{AGE50-54} + \beta_3\ \text{AGE55-59} + \beta_4\ \text{AGE60-64} \\ &+ \beta_5\ \text{HTN x AGE50-54} + \beta_6\ \text{HTN x AGE55-59} + \beta_7\ \text{HTN x AGE60-64}\end{aligned}$$

(Equation 17–7)

This model allows the hypertension effect to vary by age. Table 17–7, for example, shows how to estimate that effect in age group 50-54:

Table 17–7. The hypertension effect in age group 50-54 years

Causal assignment	Y = log (stroke rate)
HTN=1 and AGE50-54=1	$\beta_0+\beta_1 x\mathbf{1}+\beta_2 x1+\beta_3 x0+\beta_4 x0+\beta_5 x\mathbf{1}+\beta_6 x0+\beta_7 x0$
HTN=0 and AGE50-54=1	$\beta_0+\beta_1 x\mathbf{0}+\beta_2 x1+\beta_3 x0+\beta_4 x0+\beta_5 x\mathbf{0}+\beta_6 x0+\beta_7 x0$
Effect of HTN (difference in Y)	$\boldsymbol{\beta_1} + \boldsymbol{\beta_5}$

$\beta_1 + \beta_5$ = difference in Y = difference in log (stroke rate) = log (stroke rate ratio).

Rate $\text{Ratio}_{AGE50\text{-}54} = \exp(\beta_1 + \beta_5)$

And in general, to compute any age-specific effect of hypertension, we just need to re-organize the model to highlight the property of effect modification. For instance, to estimate the hypertension effect we have computed in Table17–7, combine the terms "β_1 HTN" and "β_5 HTN x AGE50-54" as shown below:

$$\begin{aligned}\log(\text{stroke rate}) = \beta_0 & \\ & + \beta_2 \text{ AGE50-54} + \beta_3 \text{ AGE55-59} + \beta_4 \text{ AGE60-64} \\ & \quad + \beta_6 \text{ HTN x AGE55-59} + \beta_7 \text{ HTN x AGE60-64} \\ & \qquad + \mathbf{(\beta_1 + \beta_5\ AGE50\text{-}54)} \text{ HTN}\end{aligned}$$

Since in that group the variable AGE50-54 takes the value of 1, the effect of hypertension is, again, $\boldsymbol{\beta_1 + \beta_5 \times 1}$.

By similar grouping of terms, we can get the hypertension effect for all four age groups (Table17–8.) Remember: these are hypertension effects, not age effects.

Table 17–8. Age-specific rate ratios of the hypertension effect

Age group	Grouped variables	Age-specific rate ratio
45-49	HTN	$\exp(\beta_1)$
50-54	HTN; HTN x AGE50-54	$\exp(\beta_1 + \beta_5)$
55-59	HTN; HTN x AGE55-59	$\exp(\beta_1 + \beta_6)$
60-64	HTN; HTN x AGE60-64	$\exp(\beta_1 + \beta_7)$

Fortunately, `PROC GENMOD` does not require us to create dummy variables or three interaction terms. If we specify the variable AGEGROUP in the "class statement" and add the product term HTN x AGEGROUP to the "model statement", the software creates all of the above.

SAS code (fit to person-based data)

```
PROC GENMOD;
 CLASS htn agegroup;
 MODEL stroke = htn agegroup htn*agegroup / DIST=POISSON
                                            LINK=LOG
                                            OFFSET=logPY;
run;
```

Selected SAS output

```
                    The GENMOD Procedure
                     Model Information

      Data Set                WORK.ONE
```

```
Distribution            Poisson
Link Function               Log
Dependent Variable       stroke
Offset Variable           logPY
Observations Used         15712

                 Class Level Information

Class          Levels    Values

htn                 2    Hypertensive Normotensive
AGEGROUP            4    A. 60-64 B. 55-59 C. 50-54 D. 45-49

Algorithm converged.


                 Analysis Of Parameter Estimates

                                                                  Standard
Parameter                                    DF        Estimate       Error

Intercept                                     1   β0  -7.8130       0.2774
htn            Hypertensive                   1   β1   2.2339       0.3203
htn            Normotensive                   0        0.0000       0.0000
AGEGROUP       A. 60-64                       1   β4   1.7536       0.3165
AGEGROUP       B. 55-59                       1   β3   1.5363       0.3157
AGEGROUP       C. 50-54                       1   β2   0.7503       0.3444
AGEGROUP       D. 45-49                       0        0.0000       0.0000
htn*AGEGROUP   Hypertensive  A. 60-64         1   β7  -1.1243       0.3675
htn*AGEGROUP   Hypertensive  B. 55-59         1   β6  -1.1902       0.3723
htn*AGEGROUP   Hypertensive  C. 50-54         1   β5  -0.8115       0.4080
htn*AGEGROUP   Hypertensive  D. 45-49         0        0.0000       0.0000
htn*AGEGROUP   Normotensive  A. 60-64         0        0.0000       0.0000
htn*AGEGROUP   Normotensive  B. 55-59         0        0.0000       0.0000
htn*AGEGROUP   Normotensive  C. 50-54         0        0.0000       0.0000
htn*AGEGROUP   Normotensive  D. 45-49         0        0.0000       0.0000
```

To the left of each estimate, I added our notation of the regression equation (equation 17–7.) After entering the estimates into Table 17–8, we get the age-specific rate ratios of the hypertension effect (Table 17–9).

Table 17–9. Age-specific rate ratios of the hypertension effect

Age group	Grouped variables	Age-specific rate ratio
45-49	HTN	$\exp(\beta_1)$ = exp(2.2339) = 9.3
50-54	HTN; HTN x AGE50-54	$\exp(\beta_1 + \beta_5)$ = exp[2.2339+(–0.8115)]=exp(1.4224)= 4.1
55-59	HTN; HTN x AGE55-59	$\exp(\beta_1 + \beta_6)$ = exp[2.2339+(–1.1902)]=exp(1.0437)= 2.8
60-64	HTN; HTN x AGE60-64	$\exp(\beta_1 + \beta_7)$ = exp[2.2339+(–1.1243)]=exp(1.1096)= 3.0

The results are identical to those we computed by tabular methods (Table 17–1). Indeed, in this example Poisson regression added nothing but complexity. Of course, if we had to condition on several confounders while estimating the modified rate ratio, tabular methods could not have delivered the goods.

One more task is still ahead: computing the standard error for each age-specific log(rate ratio) of the hypertension effect. In notation, the following standard errors are needed:

$SE[\log(RR_{AGE\ 45\text{-}49})] = SE(\beta_1)$
$SE[\log(RR_{AGE\ 50\text{-}54})] = SE(\beta_1 + \beta_5)$
$SE[\log(RR_{AGE\ 55\text{-}59})] = SE(\beta_1 + \beta_6)$
$SE[\log(RR_{AGE\ 60\text{-}64})] = SE(\beta_1 + \beta_7)$

Variance arithmetic tells us that we can't just add two standard errors to get the standard error of the sum of two coefficients. The math is a bit more complex and requires something called "covariance", which may be requested in SAS. Fortunately, however, it is possible to get the standard errors of interest by specifying the interaction model differently. The alternative code (shown below) also saves us the trouble of summing coefficients. What we get on the printout is precisely what we need: four age-specific log(rate ratios) and their standard errors.

SAS code

```
PROC GENMOD;
 CLASS agegroup htn;
 MODEL stroke = agegroup htn(agegroup) / DIST=POISSON
                                        LINK=LOG
                                        OFFSET=logPY;

run;
```

Selected SAS printout

```
                            The GENMOD Procedure

                              Model Information

          Data Set              WORK.ONE
          Distribution           Poisson
          Link Function              Log
          Dependent Variable      stroke
          Offset Variable          logPY
          Observations Used        15712

                          Class Level Information

         Class         Levels    Values
```

```
       AGEGROUP          4    A. 60-64 B. 55-59 C. 50-54 D. 45-49
       htn               2    Hypertensive Normotensive

Algorithm converged.


                      Analysis Of Parameter Estimates

                                                                 Standard
      Parameter                                 DF    Estimate       Error

      Intercept                                  1     -7.8130      0.2774
      AGEGROUP        A. 60-64                   1      1.7536      0.3165
      AGEGROUP        B. 55-59                   1      1.5363      0.3157
      AGEGROUP        C. 50-54                   1      0.7503      0.3444
      AGEGROUP        D. 45-49                   0      0.0000      0.0000
      htn(AGEGROUP)   Hypertensive  A. 60-64     1      1.1096      0.1803
      htn(AGEGROUP)   Normotensive  A. 60-64     0      0.0000      0.0000
      htn(AGEGROUP)   Hypertensive  B. 55-59     1      1.0437      0.1899
      htn(AGEGROUP)   Normotensive  B. 55-59     0      0.0000      0.0000
      htn(AGEGROUP)   Hypertensive  C. 50-54     1      1.4225      0.2528
      htn(AGEGROUP)   Normotensive  C. 50-54     0      0.0000      0.0000
      htn(AGEGROUP)   Hypertensive  D. 45-49     1      2.2339      0.3203
      htn(AGEGROUP)   Normotensive  D. 45-49     0      0.0000      0.0000
```

Table 17–10 shows the meaning of each coefficient that I highlighted on the printout, as well as the computation of 95% confidence intervals (CI) and 95% confidence limit ratios (CLR).

Table 17–10. Age-specific rate ratios of the hypertension effect and confidence intervals

Age group	β=log($RR_{AGE\text{-}SPECIFIC}$)	SE(β)	$RR_{AGE\text{-}SPECIFIC}$	95% CI*	95% CLR
45-49	2.2339	0.3203	exp(2.2339)=9.3	[5, 17]	17/5=3.4
50-54	1.4225	0.2528	exp(1.4225)=4.1	[2.5, 6.8]	6.8/2.5=2.7
55-59	1.0437	0.1899	exp(1.0437)=2.8	[2.0, 4.1]	4.1/2.0=2.1
60-64	1.1096	0.1803	exp(1.1096)=3.0	[2.1, 4.3]	4.3/2.1=2.0

* exp [$\beta \pm 1.96 SE(\beta)$]

You might have wondered why I didn't get the age-specific rate ratios by simply fitting the original regression model four times—one model per each age group—rather than by modeling interaction terms. For example, why didn't I use the following SAS code of stratified regression?

```
PROC GENMOD;
 CLASS htn;
```

```
 MODEL stroke = htn / DIST=POISSON
                      LINK=LOG
                      OFFSET=logPY;
 BY agegroup;
run;
```

Well, I could have used this code, and we would have seen identical results. Nonetheless, the two methods will often produce different estimates when the model contains covariates—for example, if we had to condition on smoking status while estimating the age-specific rate ratios of the hypertension effect. So which method should you choose in the presence of covariates: a single model that contains interaction terms or stratum-specific models?

I have raised this question before in the context of other regression models (chapter 10, for example) and answer it again here. I prefer an interaction model to stratified regression—for a reason that has nothing to do with testing a null hypothesis about the coefficients of interaction terms. When we search for modification of the hypertension effect by age group while conditioning on another variable (say, smoking status), each age-specific estimate of the hypertension effect behaves like a weighted average across the strata of smoking. If each age group contains a unique distribution of smoking status, the age-specific estimates from stratified regression will be based on different sets of weights. In contrast, every age-specific estimate from an interaction model will rely on the distribution of smoking status in the *entire* cohort—on the same set of weights.

Apart from a special likelihood function and some features of the `PROC GENMOD` code, you may think about Poisson regression along the general principles of any regression of the form "$R = \beta_0 + \beta_1 E + \ldots$" The closest analogy may be logistic regression. In a logistic model, R was "log(odds)", whereas in Poisson, R is "log(rate)". That's about it.
Poisson regression is useful and elegant, but almost nobody uses it to estimate the rate ratio unless the data file contains only group data. For person-based data from a cohort study, everyone turns to Cox regression—the next chapter.