Lesson 7
Poisson Regression

BIS 505b

Yale University
Department of Biostatistics

Pagano Chapter 7.3

Date Modified: 04/11/2021

Goals for this Lesson

Addressing a Research Question

1. How to describe count data
2. How to investigate the association between risk factors and a count outcome
   - Summarize the mean count
   - Summarize the rate when follow-up time varies
3. Identify and address overdispersion

Contents

Progress this Unit

1. Count Data
   - Poisson Distribution
   - Summarizing Count Data
2. Modeling the Count
   - Interpretation of the Model
   - Inference
   - Multiple Poisson Regression Model of the Count
3. Modeling the Rate
   - Interpretation of the Model
   - Inference and Multiple Poisson Regression of the Rate
   - Overdispersion

Count Data

- A count response Y consists of non-negative integer data, whole numbers from 0 to infinity
- Discrete, quantitative data
- Count data can be found in many public health applications
  - Number of hospitalizations
  - Number of concussions
  - Number of dental caries
  - Number of polyps removed
  - Number of relapses

Poisson Distribution

- Poisson distribution is a discrete distribution that can be used to describe the probability of observing a given number (count) of events in a fixed time interval, $Y \sim Poi(\mu)$

Poisson Probability Mass Function

$$P(Y = y \,|\, \mu) = \frac{\mu^y \, e^{-\mu}}{y!}$$

where y is an integer ≥ 0 (0, 1, 2, . . .) and $\mu > 0$

  - For a Poisson distribution, the mean and the variance of the count are equal $(E(Y) = Var(Y) = \mu)$
  - Interested in estimating μ: The mean of Y or the expected number of events during the interval

Public Health Application: Fractures and Osteoporosis

Public Health Application

Fractures of the spine, hip, and wrist are common osteoporotic fractures. Results of large prospective studies have shown that almost all types of fractures are increased in patients with low bone density and, irrespective of type of fracture, adults who sustain a fracture are 50-100% more likely to have another one of a different type.

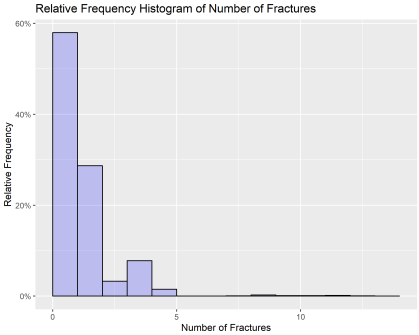
- Osteoporotic fractures are one of the most common causes of disability and a major contributor to medical care costs

A longitudinal study of postmenopausal women newly diagnosed with osteoporosis ($n = 6060$) was conducted. Women were given the opportunity to enroll in a strength training program; 3305 elected to participate. Women were contacted every 6 months to count number of new fractures. Goals of the study were to investigate:

- The impact of strength training on fracture rates, and hospitalization/urgent care rates (health care utilization rate)
- Potential risk factors related to fracture rates in this population

Summarizing Count Data

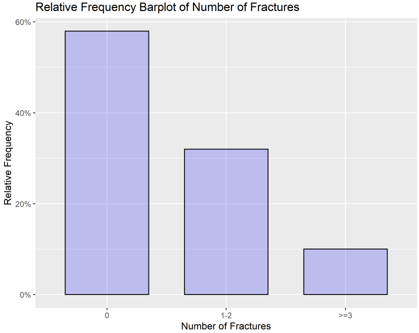
- Number of fractures during the study period in this group of newly diagnosed women can be described with summary statistics such as means, medians, standard deviations and histograms



Number of Fractures	
Mean (SD)	0.72 (1.24)
Variance	1.54
Median (Q_1, Q_3)	0 (0, 1)
Minimum, maximum	0, 13

Summarizing Count Data

- A categorical variable can be created by grouping women with similar numbers of fractures



Number of Fractures	n	%
None	3,512	57.95%
1-2	1,939	32.00%
≥ 3	609	10.05%

Progress this Unit

Regression Models

Y Variable	$E(Y\|x)$	Outcome Summarized in Terms of	Regression Model
Continuous	μ	μ	Linear Regression
Dichotomous	p	$\log\left(\frac{p}{1-p}\right)$	Logistic Regression
Count	μ	$\log(\mu)$ or $\log(\lambda)$	Poisson Regression

- Linear regression assumes the distribution of the response for a given value of x is normally distributed
- Logistic regression assumes number of observed successes follows a binomial distribution
- Poisson regression assumes the observed counts are generated from a Poisson distribution
    - Interested in the relationship between regressors and expected count

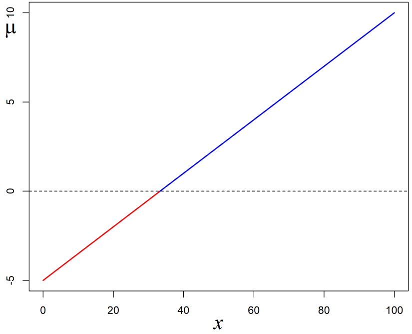
Modeling Strategy 1

- Fit a model of the form:

$$\boxed{\mu = \alpha + \beta\, x}$$

  - This is a standard linear regression model

- Problem: Even though a count response is numeric, μ is restricted to be ≥ 0. Linear regression may give negative values for $\alpha + \beta\, x$.
- Often unreasonable to assume Normality
- For a Poisson random variable, $E(Y) = Var(Y)$, which is a violation of the constant variance assumption



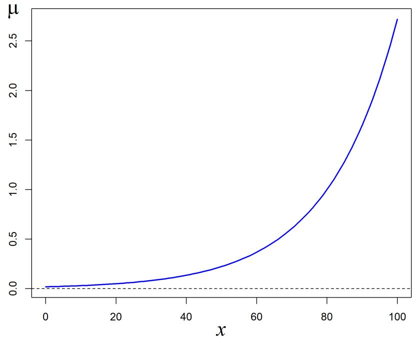
Modeling Strategy 2

- Fit a model of the form:

$$\boxed{\mu = e^{\alpha+\beta\, x}}$$

  - This is a log-linear model

- μ is guaranteed to be positive



Simple Poisson Regression Model of the Count

- Taking the (natural) log of both sides yields the Poisson regression model of the **count**:

Mean Count

$$\mu = e^{\alpha+\beta\,x}$$

Simple Poisson Regression Model of the Count

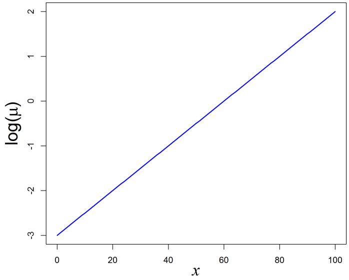
$$\log(\mu) = \alpha + \beta\,x$$

- Assumes the log of the Poisson count's expected value can be modeled by a linear combination of unknown parameters

Count response variable $\rightarrow$ Poisson regression

Poisson Regression

- Resembles a linear regression model: $\boxed{\log(\mu) = \alpha + \beta\, x}$
- Poisson regression is a generalized linear model (glm)
- Mean of Y (i.e., μ) is "linked" to $\alpha + \beta\, x$ through the log function



- Instead of assuming the relationship between μ and x is linear (Strategy 1), we assume the relationship between $\log(\mu)$ and x is linear
- Log-linear model

Interpretation of β: Change in Log Mean

- Can compare the log mean of the Poisson count $(\log(\mu))$ under two conditions by subtracting the equation for the log-mean under the two conditions, e.g., when

$x = x_1$ vs. x_0

  - $\log(\mu) = \alpha + \beta\, x$

$$\log(\mu_1) - \log(\mu_0) = [\alpha + \beta\, x_1] - [\alpha + \beta\, x_0]$$
$$= \alpha + \beta\, x_1 - \alpha - \beta\, x_0$$
$$= (x_1 - x_0)\, \beta$$

  - For example, a 1-unit increase in x $(x_1 - x_0 = 1)$ gives an expected difference of β in the log mean

$$\log(\mu_1) - \log(\mu_0) = \beta$$

Interpretation of β and e^β

$$\beta = \log(\mu_1) - \log(\mu_0) = \log\left(\frac{\mu_1}{\mu_0}\right)$$

Slope Parameter from Poisson Regression Model of the Count

Slope β from a Poisson regression model of the count is the log ratio of means associated with a 1-unit increase in risk factor x

$$e^\beta = \frac{\mu_1}{\mu_0} = \text{Mean Ratio}$$

e^β from Poisson Regression Model of the Count

e^β from a Poisson regression model of the count is the ratio of means associated with a 1-unit increase in risk factor x

Multiplicative Effect

- A change in x has a multiplicative effect on the **mean** (μ: mean count)

$$\mu_1 = \mu_0 \times e^{\beta}$$

- If $\beta = 0$, then $e^{\beta} = \dfrac{\mu_1}{\mu_0} = 1$
  - $\mu_1 = \mu_0 = e^{\alpha}$; $E(Y)$ is not related to x
- If $\beta > 0$, then $e^{\beta} = \dfrac{\mu_1}{\mu_0} > 1$ $\quad$ μ_1 is $100 \times (e^{\beta} - 1)\%$ larger than μ_0
- If $\beta < 0$, then $e^{\beta} = \dfrac{\mu_1}{\mu_0} < 1$ $\quad$ μ_1 is $100 \times (1 - e^{\beta})\%$ smaller than μ_0

Maximum Likelihood Estimation

- Maximum likelihood estimation is used to find estimates of the population parameters, α, β

$$\log(\hat{\mu}) = a + b\,x$$

- a: Estimated log mean of Y when $x = 0$
- b: Estimated log mean ratio for a 1-unit increase in x; exponentiate to give mean ratio (ratio of means)

Interpretation of Estimated Slope

$$\log(\widehat{\text{Mean Ratio}}) = b \qquad \widehat{\text{Mean Rato}} = e^{b}$$

Binary Predictor

- Consider the model containing a single binary predictor (z)

$$\log(\hat{\mu}) = a + b\,z \qquad \text{where } z = \begin{cases} 1 & \text{if Exposed} \\ 0 & \text{if Unexposed} \end{cases}$$

- $\log(\widehat{\text{Mean Ratio}})$ and $\widehat{\text{Mean Ratio}}$ of the response for exposed vs. unexposed are:

$$\log(\hat{\text{MR}}) = b \qquad \hat{\text{MR}} = e^{b}$$

- Example: Modeling mean number of fractures using treatment group

$$z = \begin{cases} 1 & \text{if strength training group} \\ 0 & \text{if no strength training} \end{cases}$$

Binary Predictor: Example

R Code, Poisson Regression of the Count

```
# Checking reference category of x, Control=reference
> contrasts(frac$group_factor)
         Strength
Control         0
Strength        1
# Fitting poisson regression model of the count
> mod1 <- glm(fractures ~ group_factor, data = frac, family=poisson(link="log"))
> summary(mod1)
Coefficients:
                     Estimate Std. Error z value             Pr(>|z|)
(Intercept)          -0.05522    0.01959  -2.819              0.00481
group_factorStrength -0.58175    0.03091 -18.818 < 0.0000000000000002
```

$$\boxed{\log(\hat{\mu}) = -0.055 - 0.582\,\text{StrengthGroup}}$$

Binary Predictor: Example

- For the data examining the relationship between number of fractures and training group, the **log expected count** of fractures is:

$$\boxed{\log(\hat{\mu}) = -0.055 - 0.582\,\textsf{StrengthGroup}}$$

R Code, Poisson Regression of the Count

```
# "a" and "b"
> coef(mod1)
        (Intercept) group_factorStrength
        -0.05521731          -0.58174689
# exp(a) and exp(b)
> exp(coef(mod1))
        (Intercept) group_factorStrength
          0.9462795            0.5589211
```

- Log ratio of mean fractures between the strength training and control group $= -0.582$

$$\log\left(\frac{\hat{\mu}_1}{\hat{\mu}_0}\right) = -0.582$$

- Estimated mean ratio: $\dfrac{\hat{\mu}_1}{\hat{\mu}_0} = e^{-0.582} = 0.56$
  - On average, women in the strength training group experienced 44% *fewer* fractures than those in the control group

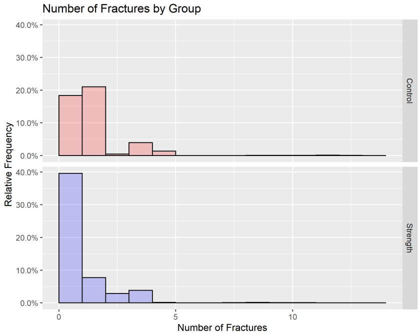
Poisson Regression of Count, Ratio of Sample Means

- As you would expect, in this unadjusted analysis, the ratio of mean fracture counts in the two groups is the ratio of group sample means

R Code, Mean Counts

```
# Mean fractures by group
> aggregate(x = list(fractures = frac$fractures),
            by = list(group = frac$group_factor),
            FUN = mean, na.rm = TRUE)
     group fractures
1  Control 0.9462795
2 Strength 0.5288956
```

$$e^b = \frac{\hat{\mu}_1}{\hat{\mu}_0} = e^{-0.582} = 0.56 = \frac{\bar{x}_1}{\bar{x}_0} = \frac{0.5289}{0.9463}$$



Poisson Regression of Count, Fitted Mean Counts

- And the fitted mean $\hat{\mu}$ number of fractures per group from the regression model equal the observed average number of falls in each group
- Control: $\log(\hat{\mu}) = -0.055 - 0.582(0) = -0.055 \qquad \hat{\mu} = e^{-0.055} = \boxed{0.9463}$
- Strength training: $\log(\hat{\mu}) = -0.055 - 0.582(1) = -0.637 \qquad \hat{\mu} = e^{-0.637} = \boxed{0.5289}$

R Code, Fitted Mean Counts

```
# Fitted mean mu_hat by group
> pred.x <- data.frame(group_factor = levels(frac$group_factor))
> cbind(pred.x, muhat = predict(mod1, newdata = pred.x, type = "response"))
  group_factor     muhat
1      Control 0.9462795
2     Strength 0.5288956
```

Hypothesis Test for β

$$\boxed{\log(\mu) = \alpha + \beta x}$$

- The next step is to determine if there is a significant relationship between x and the **mean** Poisson count
- **Hypothesis test** for the slope parameter β (Wald Test)
  - H_0: $\beta = 0$, equivalent to $\dfrac{\mu_1}{\mu_0} = 1$ or $\mu_1 = \mu_0$
  - H_1: $\beta \neq 0$, equivalent to $\dfrac{\mu_1}{\mu_0} \neq 1$ or $\mu_1 \neq \mu_0$

Wald Test Statistic for Slope

$$Z = \frac{b}{s_b} \sim N(0,1)$$

Hypothesis Test for β: Example

- Example: Is there a difference in **mean** fracture count in strength training vs. control group? H_0: $\beta = 0$ ($\mu_1 = \mu_0$) vs. H_1: $\beta \neq 0$

R Code, Poisson Regression of the Count

```
> mod1 <- glm(fractures ~ group_factor, data = frac, family=poisson(link="log"))
> summary(mod1)
Coefficients:
                     Estimate Std. Error z value              Pr(>|z|)
(Intercept)          -0.05522    0.01959  -2.819               0.00481
group_factorStrength -0.58175    0.03091 -18.818 < 0.0000000000000002
```

- $z = \dfrac{b}{s_b} = -\dfrac{0.5818}{0.0309} = -18.82$ compared to $N(0,1)$
  - Reject H_0 if $|z| \geq z_{1-\frac{\alpha}{2}} = z_{.975} = z^* = 1.96$
  - $|z| = 18.82 \geq z^* \rightarrow$ Reject H_0

```
pval = 2*(1-pnorm(18.82))
```

  - $p = 2 \times P(Z \geq 18.82) < .0001$
  - $p < 0.05 \rightarrow$ Reject H_0

- Conclusion: There is a significant difference in mean fracture count between the strength training and placebo groups ($p < .0001$)

Confidence Interval for $\log(\text{Mean Ratio})$ and Mean Ratio

- **Confidence interval** for the slope β, $\log(\mu_1/\mu_0)$, has the form:

$100(1-\alpha)\%$ Confidence Interval for $\log(\mu_1/\mu_0), \beta$
$b \pm z_{1-\frac{\alpha}{2}}\ s_b = (c_L\,, c_U\,)$

- To find the confidence interval for the mean ratio, exponentiate the lower and upper bounds of the CI for $\log(\mu_1/\mu_0)$

$100(1-\alpha)\%$ Confidence Interval for μ_1/μ_0
(e^{c_L}, e^{c_U})

- Note: The confidence interval for the mean ratio will *exclude* 1 if we rejected H_0. The CI will *include* 1 if we failed to reject H_0.

Confidence Interval for $\log$(Mean Ratio) and Mean Ratio: Example

- Example: Confidence interval of μ_1/μ_0 in strength training vs. control

R Code, Confidence Intervals

```
> confint.default(mod1)  # CI for beta (log MR)
                              2.5 %      97.5 %
(Intercept)             -0.09360372 -0.01683089
group_factorStrength -0.64233686 -0.52115692
> exp(cbind(MR = coef(mod1), confint.default(mod1)))  # MR and CI for MR
                             MR     2.5 %    97.5 %
(Intercept)             0.9462795 0.9106436 0.9833100
group_factorStrength 0.5589211 0.5260617 0.5938331
```

- 95% CI for $\log$(mean ratio), β:

$$b \pm z_{1-\frac{\alpha}{2}}\, s_b = -0.5818 \pm 1.96 \times 0.0309 = (c_L = -0.6423, c_U = -0.5212)$$

- 95% CI for mean ratio: $(e^{-0.6423}, e^{-0.5212}) = (0.5261, 0.5938)$, which excludes 1

Multiple Poisson Regression Model of the Count: Example

- Because women in this study self-selected to participate in strength training, possible that confounding factors bias the comparison of mean fracture counts between women participating in strength-training or not
- Have information on age, race, calcium use, history of fractures, quality of life, and pain scores
- Question: Is strength training associated with the **average number** of fractures, while controlling for other participant characteristics?

Characteristics of Study Cohort by Group: Example

Table 1: Summary Statistics of Fracture Study Cohort

Intervention Group	Control (N=2755)	Strength (N=3305)	Total (N=6060)
Number of Fractures			
Mean (SD)	0.9 (1.3)	0.5 (1.1)	0.7 (1.2)
Median (Range)	1.0 (0.0, 13.0)	0.0 (0.0, 11.0)	0.0 (0.0, 13.0)
Age (years)			
Mean (SD)	60.7 (9.2)	56.9 (9.7)	58.6 (9.7)
Median (Range)	61.0 (39.0, 85.0)	56.0 (31.0, 85.0)	58.0 (31.0, 85.0)
Pain Score			
Mean (SD)	5.9 (2.2)	5.5 (2.7)	5.7 (2.5)
Median (Range)	6.0 (0.0, 10.0)	6.0 (0.0, 10.0)	6.0 (0.0, 10.0)
Quality of Life Score			
Mean (SD)	44.5 (15.4)	44.8 (14.9)	44.7 (15.2)
Median (Range)	45.0 (0.0, 97.0)	45.0 (0.0, 96.0)	45.0 (0.0, 97.0)
Calcium Use			
No	691 (25.1%)	849 (25.7%)	1540 (25.4%)
Yes	2064 (74.9%)	2456 (74.3%)	4520 (74.6%)
Fracture History			
No	1288 (46.8%)	1600 (48.4%)	2888 (47.7%)
Yes	1467 (53.2%)	1705 (51.6%)	3172 (52.3%)
Race			
White	1804 (65.5%)	2184 (66.1%)	3988 (65.8%)
Black	542 (19.7%)	657 (19.9%)	1199 (19.8%)
Other	409 (14.8%)	464 (14.0%)	873 (14.4%)

Multiple Poisson Regression Model of the Count: Example

R Code, Poisson Regression of the Count

```
> mod2 <- glm(fractures ~ group_factor + age + pain + qol + calcium_factor + frachistory_factor +
              race_factor, data = frac, family=poisson(link="log"))
> summary(mod2)
Coefficients:
                        Estimate Std. Error z value              Pr(>|z|)
(Intercept)           -0.9952576  0.1227215  -8.110 0.000000000000000507
group_factorStrength  -0.5438336  0.0316011 -17.209 < 0.0000000000000002
age                    0.0094698  0.0015994   5.921 0.000000003204526810
pain                  -0.0154570  0.0061589  -2.510               0.0121
qol                    0.0024014  0.0009927   2.419               0.0156
calcium_factorYes     -0.3956289  0.0321617 -12.301 < 0.0000000000000002
frachistory_factorYes  1.0099327  0.0351204  28.756 < 0.0000000000000002
race_factorBlack      -0.0797605  0.0399205  -1.998               0.0457
race_factorOther      -0.1017656  0.0452305  -2.250               0.0245
```

$$\log(\hat{\mu}) = -1.00 - 0.54\,\text{StrengthGroup} + 0.009\,\text{Age} - 0.015\,\text{Pain} + 0.0024\,\text{QOL} - 0.40\,\text{Calcium} + 1.01\,\text{FracHistory} - 0.08\,\text{BlackRace} - 0.10\,\text{OtherRace}$$

Multiple Poisson Regression Model of the Count: Example

R Code, Confidence Intervals

```
# OR and CI for OR
> exp(cbind(MR = coef(mod2), confint.default(mod2)))
                              MR     2.5 %    97.5 %
(Intercept)            0.3696282 0.2906059 0.4701385
group_factorStrength   0.5805185 0.5456538 0.6176109
age                    1.0095148 1.0063551 1.0126843
pain                   0.9846618 0.9728471 0.9966200
qol                    1.0024043 1.0004558 1.0043566
calcium_factorYes      0.6732565 0.6321272 0.7170619
frachistory_factorYes  2.7454164 2.5627939 2.9410524
race_factorBlack       0.9233375 0.8538471 0.9984833
race_factorOther       0.9032413 0.8266154 0.9869703
```

- Just as in logistic regression, a likelihood ratio test can be used to simultaneously test multiple β
  - H_0: $\beta_7 = \beta_8 = 0$
  - H_1: β_7 and β_8 are not both 0

R Code, LRT

```
# Load required package
> library(car)
> Anova(mod2)
Analysis of Deviance Table (Type II tests)

Response: fractures
                   LR Chisq Df        Pr(>Chisq)
group_factor         303.15  1 < 0.000000000000022
age                   34.87  1      0.0000003525
pain                   6.27  1           0.01228
qol                    5.85  1           0.01559
calcium_factor       144.11  1 < 0.000000000000022
frachistory_factor   949.38  1 < 0.000000000000022
race_factor            7.73  2           0.02100
```

- Race: $G = 7.73$ compared to χ^2_2 ($p = 0.02$) $\rightarrow$ Race is a significant predictor in this model

Multiple Poisson Regression Model of the Count: Example

Parameter	Estimate	Mean Ratio ($\exp(b_j)$)	p-value
Intercept	-0.9953		<.0001
Group (Strength vs. None)	-0.5438	0.581	<.0001
Age	0.0095	1.010	<.0001
Pain	-0.0155	0.985	0.0121
QOL	0.0024	1.002	0.0156
Calcium (Y vs. N)	-0.3956	0.673	<.0001
FracHistory (Y vs. N)	1.0099	2.745	<.0001
Race (B vs. W)	-0.0798	0.923	0.0457
Race (O vs. W)	-0.1018	0.903	0.0245

- Even after controlling for potential confounders, strength training reduces mean number of fractures by 42%
- A 1-year increase in age increases the expected number of fractures by 1% (5-year increase: $e^{5\times 0.0095} = 1.05$)
- An increase in pain score, calcium consumption associated with decrease in expected number of fractures
- Increasing QOL score, history of fractures associated with increase in expected number of fractures
- Black women experience 7.7% fewer fractures than white women, on average
- Women of other races experience 9.7% fewer fractures than white women, on average

Progress this Unit

Accounting for Time at Risk

- Count data are usually collected over a period of time
- Consequently, the length of follow-up may vary from subject to subject
- Our previous analysis of the mean fracture count assumes each woman contributes equally to the count (i.e., that each subject is followed for the same amount of time)
  - Treating a woman who has been observed for 5 years the same as a woman who has been observed for only 1 year is unreasonable
  - The woman observed for 5 years has more opportunities to have fractures and may be expected to have a larger fracture count than a woman who had only 1 year of follow-up

Summarizing Count Data: Time at Risk

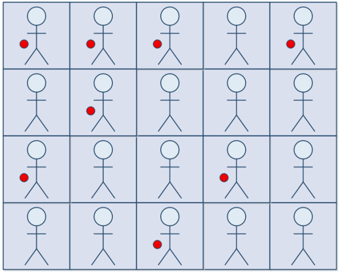
Table 2: Summary Statistics of Follow-up Time in Study Cohort

Intervention Group	Control (N=2755)	Strength (N=3305)	Total (N=6060)
Number of Fractures			
Mean (SD)	0.95 (1.33)	0.53 (1.13)	0.72 (1.24)
Median (Q1, Q3)	1.00 (0.00, 1.00)	0.00 (0.00, 1.00)	0.00 (0.00, 1.00)
Min - Max	0.00 - 13.00	0.00 - 11.00	0.00 - 13.00
Follow-up Time (years)			
Mean (SD)	6.98 (1.18)	7.05 (1.19)	7.02 (1.18)
Median (Q1, Q3)	6.98 (6.21, 7.75)	7.05 (6.25, 7.85)	7.02 (6.23, 7.80)
Min - Max	2.96 - 11.24	2.74 - 11.08	2.74 - 11.24

- Thus, investigations of count data also need to consider potential variations in the contribution of time at risk for an event, when appropriate
- To account for differing amounts of observation for each subject, count outcomes are often described with event rates

Event Rate vs. Proportion

Figure: Visualization of Calculating a Proportion

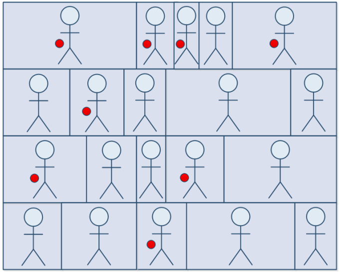


Proportion: (unitless)

- Numerator: Number of subjects with the event
- Denominator: Total number of subjects
- Each subject contributes the same amount to the denominator and at most one event to the numerator
- Denominator represents all subjects, giving each subject an equal weight (each subject has the same size square)

Event Rate vs. Proportion

Figure: Visualization of Calculating an Event Rate



Event Rate: (events/person-time)

- Numerator: Number of events (can be multiple)
- Denominator: Total amount of time at risk for event
- Each subject not treated equally in the denominator: Some subjects contribute bigger squares because they are followed longer

Event Rate

- λ: Rate of events in the population
- $\hat{\lambda}$: Event rate is estimated in the sample by dividing the total number of events by the accumulated person-time in the study (person-years, person-months, person-days)

Sample Event Rate

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} t_i}$$

Where: y_i: Number of events for subject i
t_i: Amount of time subject i is at risk

Event Rate: Example

Fractures	t_i (years)
1	9.14653
1	7.18186
0	6.70453
1	6.36389
0	6.93181
0	7.02789
1	9.73184
0	6.79541
3	6.39919
0	7.75024
0	9.03693
1	8.76396
2	9.48711
$\vdots$	
$\sum y_i = 4355$	$\sum t_i = 42535.6$

- Total number of fractures = 4355
- Total number of years of follow-up for fractures = 42,535.6 person-years

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} t_i} = \frac{4355 \text{ fractures}}{42535.6 \text{ person-years}} = 0.1 \text{ fractures/person-year}$$

Rate Ratio

- In the study of fractures, the primary outcome is a count, the number of fractures
- However, because the amount of time the subjects are observed varies, the fracture rate becomes of primary interest
- The research question is whether strength training has an impact on fracture rates and can be answered by estimating how the fracture rate in the strength training group compares to the fracture rate in the control group, or the rate ratio

Rate Ratio

$$\widehat{\text{RR}} = \frac{\text{Event rate in exposed}}{\text{Event rate in unexposed}} = \frac{\hat{\lambda}_1}{\hat{\lambda}_0}$$

Rate Ratio: Example

Group	Events	Person-yrs	$\hat{\lambda}^*$
Strength training group	1748	23304.63	0.075
Control group	2607	19231	0.1356

*cases/person-year

$$\widehat{\text{RR}} = \frac{\hat{\lambda}_1}{\hat{\lambda}_0} = \frac{0.075}{0.136} = 0.55$$

- On average, women in the strength training group experienced a 45% lower fracture rate $(100 \times (1 - 0.55)\%)$ than women in the control group

R Code, Crude Rates by Group

```
> bygroup.sum <- aggregate(x = list(fractures = frac$fractures, time = frac$time),
                           by = list(group = frac$group_factor), FUN = sum, na.rm = TRUE)
> bygroup.rate <- cbind(bygroup.sum, rate = bygroup.sum[,2]/bygroup.sum[,3])
> bygroup.rate
     group fractures     time       rate
1  Control      2607 19231.00 0.13556234
2 Strength      1748 23304.63 0.07500656
```

Time Scale

- Re-scaling the observed follow-up time will not affect the estimated rate ratio
- For example:

$$\widehat{\text{RR}} = 2 = \frac{\dfrac{10 \text{ events}}{\text{person-year}}}{\dfrac{5 \text{ events}}{\text{person-year}}} = \frac{\dfrac{10 \text{ events}}{12 \text{ person-months}}}{\dfrac{5 \text{ events}}{12 \text{ person-months}}}$$

- Will affect the scale of the individual rate (10 events per person-year is equivalent to 0.833 events per person-month)

Poisson Regression Model for Rates

- When outcomes occur over time, more relevant to model their rate of occurrence than the raw count
- When a response count, y, has time at risk associated with it equal to t, the sample rate is $\hat{\lambda} = y/t$. Its expected value is μ/t.
- The **rate** of the event, is often modeled as a log-linear model since the rate λ is constrained to be positive, but does not have to be in the interval $[0, 1]$

Rate

$$\lambda = e^{\alpha + \beta\, x}$$

Simple Poisson Regression Model of the Rate

$$\log(\lambda) = \log\left(\frac{\mu}{t}\right) = \alpha + \beta\, x$$

$\boxed{\text{Rate response variable} \rightarrow \text{Poisson regression}}$

Poisson Regression Model for Rates

- Poisson regression model of the expected **rate** of the event:

$$\begin{aligned} \log(\lambda) = \log\left(\frac{\mu}{t}\right) &= \alpha + \beta\, x \\ \log(\mu) - \log(t) &= \alpha + \beta\, x \\ \log(\mu) &= \alpha + \beta\, x + \overset{\text{offset}}{\boxed{\log(t)}} \end{aligned}$$

- Reduces to a model for the Poisson **count** with an adjustment term ("offset"), $\log(t)$, that accounts for each individual's time at risk
- Notice that there is no parameter for the offset term ("offset" is a term used for a predictor whose β is forced to be equal to one)

Simple Poisson Regression Model of the Rate

$$\log(\mu) = \alpha + \beta\, x + \log(t)$$

Interpretation of β: Change in Log Rate

- Can compare the log rate ($\log(\lambda)$) under two conditions, e.g., $x = x_1$ vs. x_0
  - $\log(\lambda) = \alpha + \beta\, x$

$$\begin{aligned}\log(\lambda_1) - \log(\lambda_0) &= [\alpha + \beta\, x_1] - [\alpha + \beta\, x_0] \\ &= \alpha + \beta\, x_1 - \alpha - \beta\, x_0 \\ &= (x_1 - x_0)\,\beta\end{aligned}$$

  - For example, a 1-unit increase in x ($x_1 - x_0 = 1$) gives an expected difference of β in the log rate

$$\log(\lambda_1) - \log(\lambda_0) = \beta$$

Interpretation of β and e^{β}

$$\beta = \log(\lambda_1) - \log(\lambda_0) = \log\left(\frac{\lambda_1}{\lambda_0}\right)$$

Slope Parameter from Poisson Regression Model of the Rate

Slope β from a Poisson regression model of the rate is the log rate ratio associated with a 1-unit increase in risk factor x

$$e^{\beta} = \frac{\lambda_1}{\lambda_0} = \text{RR}$$

e^{β} from Poisson Regression Model of the Rate

e^{β} from a Poisson regression model of the rate is the rate ratio associated with a 1-unit increase in risk factor x

Multiplicative Effect

- A change in x has a multiplicative effect on the **rate**

$$\lambda_1 = \lambda_0 \times e^{\beta}$$

- If $\beta = 0$, then $e^{\beta} = \frac{\lambda_1}{\lambda_0} = 1$
  - $\lambda_1 = \lambda_0 = e^{\alpha}$; the rate of the event is not related to x
- If $\beta > 0$, then $e^{\beta} = \frac{\lambda_1}{\lambda_0} > 1$ $\quad$ λ_1 is $100 \times (e^{\beta} - 1)\%$ larger than λ_0
- If $\beta < 0$, then $e^{\beta} = \frac{\lambda_1}{\lambda_0} < 1$ $\quad$ λ_1 is $100 \times (1 - e^{\beta})\%$ smaller than λ_0

Simple Poisson Regression Model of the Rate: Example

- Example: We now want to account for the different follow-up times of participants. Suppose we want to compare the **rate** of fractures experienced in women in the strength training group vs. the control group

$$z = \begin{cases} 1 & \text{if strength training group} \\ 0 & \text{if no strength training} \end{cases}$$

$$\log(\hat{\lambda}) = a + b\,\text{StrengthGroup}$$

- $\log(\widehat{\text{RR}})$ and $\widehat{\text{RR}}$ of fractures in strength training group vs. the control group are:

$$\log(\widehat{\text{RR}}) = b \qquad\qquad \widehat{\text{RR}} = e^{b}$$

Binary Predictor: Example

$$\boxed{\log(\mu) = \alpha + \beta\, x + \log(t)}$$

R Code, Poisson Regression of the Rate

```
> summary(frac$time)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.742   6.234   7.016   7.019   7.799  11.242
# Fitting poisson regression model of the rate
> mod3 <- glm(fractures ~ group_factor + offset(log(time)), data = frac,
              family=poisson(link="log"))
> summary(mod3)
...
Coefficients:
                     Estimate Std. Error z value             Pr(>|z|)
(Intercept)          -1.99832    0.01959 -102.03 <0.0000000000000002
group_factorStrength -0.59186    0.03091  -19.14 <0.0000000000000002
```

$$\boxed{\log(\hat{\lambda}) = -1.998 - 0.592\,\text{StrengthGroup}}$$

Binary Predictor: Example

- For the data examining the relationship between number of fractures and training group, the **log expected fracture rate** is:

$$\boxed{\log(\hat{\lambda}) = -1.998 - 0.592\,\textsf{StrengthGroup}}$$

R Code, Poisson Regression of the Rate

```
# "a" and "b"
> coef(mod3)
        (Intercept) group_factorStrength
          -1.998324            -0.591856

# exp(a) and exp(b)
> exp(coef(mod3))
        (Intercept) group_factorStrength
          0.1355623            0.5532994
```

- Log rate ratio between the strength training and control group $= -0.592$

$$\log\left(\frac{\hat{\lambda}_1}{\hat{\lambda}_0}\right) = -0.592$$

- Estimated rate ratio: $\dfrac{\hat{\lambda}_1}{\hat{\lambda}_0} = e^{-0.592} = 0.55$
  - On average, women in the strength training group experienced 45% *lower* fracture rate than the control group

Hypothesis Test for β

$$\boxed{\log(\lambda) = \alpha + \beta\, x}$$

- The next step is to determine if there is a significant relationship between x and the **rate**
- **Hypothesis test** for the slope parameter β (Wald Test)
  - H_0: $\beta = 0$, equivalent to $\dfrac{\lambda_1}{\lambda_0} = 1$ or $\lambda_1 = \lambda_0$
  - H_1: $\beta \neq 0$, equivalent to $\dfrac{\lambda_1}{\lambda_0} \neq 1$ or $\lambda_1 \neq \lambda_0$

Wald Test Statistic for Slope

$$Z = \frac{b}{s_b} \sim N(0,1)$$

Poisson Regression of Count, Fitted Rate

- The fitted yearly fracture rate $\hat{\lambda}$ per group from the regression model equal the observed fracture rate in each group
- Control: $\log(\hat{\lambda}) = -1.998 - 0.592(0) = -1.998$ $\quad \hat{\lambda} = e^{-1.998} = \boxed{0.136}$
- Strength training: $\log(\hat{\lambda}) = -1.998 - 0.592(1) = -2.59$ $\quad \hat{\lambda} = e^{-2.59} = \boxed{0.075}$

R Code, Fitted Rate

```
# Fitted rate lambda_hat by group
> pred.x <- data.frame(group_factor = levels(frac$group_factor), time = 1)
> cbind(pred.x, muhat = predict(mod3, newdata = pred.x, type = "response"))
  group_factor time  lambdahat
1      Control    1 0.13556234
2     Strength    1 0.07500656
```

Hypothesis Test for β: Example

- Example: Is there a difference in the fracture **rate** in strength training vs. control group?
  H_0: $\beta = 0$ $(\lambda_1 = \lambda_0)$ vs. H_1: $\beta \neq 0$

R Code, Poisson Regression of the Count

```
> mod3 <- glm(fractures ~ group_factor + offset(log(time)), data=frac, family=poisson(link="log"))
> summary(mod3)
Coefficients:
                    Estimate Std. Error z value              Pr(>|z|)
(Intercept)         -1.99832    0.01959 -102.03 <0.0000000000000002
group_factorStrength -0.59186    0.03091  -19.14 <0.0000000000000002
```

- $z = \dfrac{b}{s_b} = -\dfrac{0.5919}{0.0309} = -19.14$ compared to $N(0,1)$

  `pval = 2*(1-pnorm(19.14))`

  - Reject H_0 if $|z| \geq z_{1-\frac{\alpha}{2}} = z_{.975} = z^* = 1.96$
  - $|z| = 19.14 \geq z^* \rightarrow$ Reject H_0
  - $p = 2 \times P(Z \geq 19.14) < .0001$
  - $p < 0.05 \rightarrow$ Reject H_0
- Conclusion: There is a significant difference in fracture rate between the strength training and placebo groups $(p < .0001)$

Confidence Interval for $\log(\text{Rate Ratio})$ and Rate Ratio

- **Confidence interval** for the slope β, $\log(\lambda_1/\lambda_0)$, has the form:

$100(1-\alpha)\%$ Confidence Interval for $\log(\lambda_1/\lambda_0), \beta$

$$b \pm z_{1-\frac{\alpha}{2}}\ s_b = (c_L\,, c_U\,)$$

- To find the confidence interval for the rate ratio, exponentiate the lower and upper bounds of the CI for $\log(\lambda_1/\lambda_0)$

$100(1-\alpha)\%$ Confidence Interval for λ_1/λ_0

$$(e^{c_L}\,, e^{c_U}\,)$$

- Note: The confidence interval for the rate ratio will *exclude* 1 if we rejected H_0. The CI will *include* 1 if we failed to reject H_0.

Confidence Interval for $\log(\text{Rate Ratio})$ and Rate Ratio: Example

- Example: Confidence interval of λ_1/λ_0 in strength training vs. control

R Code, Confidence Intervals

```
> confint.default(mod3)  # CI for beta (log RR)
                          2.5 %    97.5 %
(Intercept)          -2.036710 -1.959937
group_factorStrength -0.652446 -0.531266
> exp(cbind(MR = coef(mod3), confint.default(mod3)))  # RR and CI for RR
                            MR     2.5 %    97.5 %
(Intercept)          0.1355623 0.1304572 0.1408673
group_factorStrength 0.5532994 0.5207704 0.5878603
```

- 95% CI for $\log(\text{rate ratio})$, β:

$$b \pm z_{1-\frac{\alpha}{2}}\, s_b = -0.5919 \pm 1.96 \times 0.0309 = (c_L = -0.6524, c_U = -0.5313)$$

- 95% CI for rate ratio: $(e^{-0.6524}, e^{-0.5313}) = (0.5208, 0.5879)$, which excludes 1

Multiple Poisson Regression Model of the Rate: Example

- Question: Is strength training associated with the **rate** of fractures, while controlling for other variables in the model?

R Code, Poisson Regression of the Rate

```
> mod4 <- glm(fractures ~ group_factor + age + pain + qol + calcium_factor + frachistory_factor +
          race_factor + offset(log(time)), data=frac, family=poisson(link="log"))
```

	Estimate	SE	RR	p-value
Intercept	-2.9527	0.123		$<.0001$
Group (Strength vs. None)	-0.5548	0.032	0.574	$<.0001$
Age	0.0097	0.002	1.010	$<.0001$
Pain	-0.0140	0.006	0.986	0.0229
QOL	0.0023	0.001	1.002	0.0229
Calcium (Y vs. N)	-0.4005	0.032	0.670	$<.0001$
FracHistory (Y vs. N)	1.0156	0.035	2.761	$<.0001$
Race (B vs. W)	-0.0768	0.040	0.926	0.0543
Race (O vs. W)	-0.1026	0.045	0.902	0.0233

- The adjusted rate ratio for the strength training group is 0.574; strength training results in a 42.6% reduction in fracture **rate**

Overdispersion

- When using Poisson regression, assumption is that the count data arose from a Poisson distribution
- One feature of the Poisson distribution is that the mean equals the variance
- If there is evidence that the variance is not similar to the mean, the Poisson distribution may not be appropriate, and the model may not be appropriate
- Overdispersion occurs when the variance is larger than expected
  - Overdispersion is not an issue in ordinary regression because a Normal distribution has a separate variance parameter, σ^2
  - In the presence of overdispersion, the parameter estimates are consistent
  - However, standard errors will be underestimated, affecting model inference (p-values and confidence intervals)

Detecting Overdispersion

- The residual deviance and its degrees of freedom are used to detect overdispersion
- Residual deviance / residual degrees of freedom $(n-(p+1)) \approx 1$ if no overdispersion is present (as rule of thumb, ratio $> 1.1 \rightarrow$ overdispersion)

Indicator of Overdispersion

$$\text{Residual Deviance/Residual df} >> 1$$

Detecting Overdispersion

R Code, Fit of Poisson Regression

```
> summary(mod4)
...
    Null deviance: 9699.9  on 6056  degrees of freedom
Residual deviance: 8167.1  on 6048  degrees of freedom
  (3 observations deleted due to missingness)
> deviance(mod4)
[1] 8167.126
> mod4$df.residual
[1] 6048
# If >> 1, indicates overdispersion
> deviance(mod4)/mod4$df.residual
[1] 1.350385
```

- Ratio of residual deviance to its degrees of freedom = 1.35, here indicating overdispersion is present

Negative Binomial Regression

- Negative binomial regression is often used as an alternative to Poisson regression for a count outcome when overdispersion is present
- The model is similar to Poisson regression model, but the distribution associated with the model is the negative binomial instead of the Poisson and does not assume $E(Y) = Var(Y)$
  - $E(Y) = \mu$
  - $Var(Y) = \mu + \dfrac{\mu^2}{k}$
  - $\dfrac{1}{k}$ is the dispersion parameter that isisson used to adjust the variance independently of the mean; equals 0 in Poisson regression

Negative Binomial Regression: Example

R Code, Negative Binomial Regression

```
> library(MASS)  # Load required package
> mod4.nb <- glm.nb(fractures ~ group_factor + age + pain + qol + calcium_factor
          + frachistory_factor + race_factor + offset(log(time)), data = frac) # Neg. bin. glm
> summary(mod4.nb)
Coefficients:
(Intercept)            -2.918908   0.155929 -18.719 < 0.0000000000000002
group_factorStrength   -0.547725   0.039839 -13.749 < 0.0000000000000002
age                     0.009852   0.002049   4.808         0.00000152
pain                   -0.013662   0.007827  -1.745             0.0809
...
(Dispersion parameter for Negative Binomial(1.3931) family taken to be 1)
> k <- mod4.nb$theta
> 1/k  # dispersion parameter (Estimate of 1/k = 1/1.3931)
[1] 0.7178152
```

- Interpretation of NB regression results the same as Poisson regression (mean ratio, rate ratio)

$$\boxed{\log(\hat{\lambda}) = -2.92 - 0.555\,\text{StrengthGroup} + 0.0097\,\text{Age} - 0.0137\,\text{Pain}\ldots}$$

Negative Binomial Regression: Example

- Recall, for NB glm, $Var(Y) = \mu + \dfrac{\mu^2}{k}$, so at an estimated mean $\hat{\mu}$, the estimated variance equals $\hat{\mu} + 0.72\hat{\mu}^2$, compared to $\hat{\mu}$ for a Poisson glm
- Fitted values are similar, but greater estimated variance in NB model and resulting greater SE for b_js reflect the overdispersion uncaptured with Poisson glm

	Poisson				Negative Binomial			
	Estimate	SE	RR	p-value	Estimate	SE	RR	p-value
Intercept	-2.9527	0.123		$<.0001$	-2.9189	0.156		$<.0001$
Group (Strength vs. None)	-0.5548	0.032	0.574	$<.0001$	-0.5477	0.040	0.578	$<.0001$
Age	0.0097	0.002	1.010	$<.0001$	0.0099	0.002	1.010	$<.0001$
Pain	-0.0140	0.006	0.986	0.0229	-0.0137	0.008	0.986	**0.0809**
QOL	0.0023	0.001	1.002	0.0229	0.0019	0.001	1.002	**0.1293**
Calcium (Y vs. N)	-0.4005	0.032	0.670	$<.0001$	-0.4196	0.042	0.657	$<.0001$
FracHistory (Y vs. N)	1.0156	0.035	2.761	$<.0001$	1.0079	0.042	2.740	$<.0001$
Race (B vs. W)	-0.0768	0.040	0.926	0.0543	-0.0900	0.051	0.914	**0.0758**
Race (O vs. W)	-0.1026	0.045	0.902	0.0233	-0.1220	0.058	0.885	**0.0340**

Testing if $\frac{1}{k} = 0$

- Likelihood Ratio Test can be used to determine if overdispersion is present and the negative binomial model provides a better fit than the Poisson model

Likelihood Ratio Test Statistic

$$
\begin{aligned}
\text{G} &= -2\log\text{-likelihood}(R) - (-2\log\text{-likelihood}(F)) \\
&= -2\left[\log\text{-likelihood}(Poisson) - \log\text{-likelihood}(NB)\right]
\end{aligned}
$$

- Compared to a χ^2_1
- $\text{G} = 14063.57 - (13338.81) = 724.758 > 3.84 = \chi_{.95,1}$
- Conclude the negative binomial model provides a better fit than the Poisson model

R Code, LRT

```
# LRT -2LL.Poisson - (-2LL.NB)
G <- as.numeric(-2*logLik(mod4)-(-2*logLik(mod4.nb)))
pval <- 1 - pchisq(G, df = 1)
```

Public Health Application: Poisson Regression of the Count

Public Health Application

Systemic effect of water fluoridation on dental caries prevalence

Cho HJ, Jin BH, Park DY, Jung SH, Lee HS, Paik DI, Bae KH. Systemic effect of water fluoridation on dental caries prevalence. Community Dent Oral Epidemiol 2014. © 2014 John Wiley & Sons A/S. Published by John Wiley & Sons Ltd

Abstract – *Objectives:* The aim of this study was to evaluate the systemic effect of water fluoridation on dental caries prevalence and experience in Cheongju, South Korea, where water fluoridation ceased 7 years previously. *Methods:* A cross-sectional survey was employed at two schools where water fluoridation had ceased (WF-ceased area) and at two schools where the water had never been fluoridated (non-WF area). The schools in the non-WF area were of a similar population size to the schools in the WF-ceased area. Children of three age groups were examined in both areas: aged 6 ($n = 505$), 8 ($n = 513$), and 11 years ($n = 467$). The differences in the mean number of decayed or filled primary teeth (dft) and the mean number of decayed, missing, or filled permanent teeth (DMFT) scores between areas after adjusting for oral health behaviors and socio-demographic factors were analyzed by a Poisson regression model.

Link to article

Journal Example: Poisson Regression Model of the Count

Public Health Application

Results The regression model showed that the DMFT ratio of the mean DMFT score for 11-year-olds in the WF-ceased area versus non-WF area was 0.581 (95% CI 0.450–0.751). Conversely, the dft ratio for 6-year-olds in the WF-ceased area versus non-WF area was 1.158 (95% CI 1.004–1.335). This showed that the adjusted mean DMFT in the WF-ceased area was 16% greater for 6-year-olds and 42% lower for 11-year-olds than in the non-WF area. The DMFT ratio for 8-year-olds (0.924, 95% CI 0.625–1.368) was not statistically significant.

- 11-year-old children in the WF-ceased area who had ingested fluoridated water for approximately 4 years after birth showed significantly lower DMFT than those in the non-WF area (estimated mean ratio = 0.581)
- This suggests that the systemic effect of fluoride intake through water fluoridation could be important for the prevention of dental caries

Public Health Application: Poisson Regression of the Rate

Public Health Application

Readmissions and the quality of care in patients hospitalized with heart failure

JEAN-CHRISTOPHE LUTHI[1,2,3], MARY JO LUND[2], LAURA SAMPIETRO-COLOM[2,3,4], DAVID G. KLEINBAUM[2], DAVID J. BALLARD[5] AND WILLIAM M. MCCLELLAN[2,6]

Abstract

Objectives. Clinical practice guidelines based on the results of randomized clinical trials recommend that patients with heart failure due to left ventricular systolic dysfunction (LVSD) be treated with angiotensin-converting enzyme inhibitors (ACEI) at doses shown to reduce mortality and readmission. This study examined the relationship between ACEI use at discharge and readmission among patients with heart failure due to LVSD.

Methods and results. Data were abstracted from the medical records of 2943 randomly selected patients hospitalized for heart failure in 50 hospitals. The outcome of interest was the number of readmissions occurring up to 21 months after discharge. Six-hundred and eleven patients were eligible for analysis. Compared with patients discharged at a recommended ACEI dose, patients not prescribed an ACEI at discharge had an adjusted rate ratio of readmission (RR) of 1.74 [95% confidence interval (CI) 1.22–2.48], while patients prescribed an ACEI at less than a recommended dose had an RR of 1.24 (95% CI 0.91–1.69) ($P = 0.005$ for the trend).

International Journal for Quality in Health Care 2003; **Volume 15, Number 5**: pp. 413–421

Link to article

Journal Example: Poisson Regression Model of the Rate

Public Health Application

Study design

This was a retrospective cohort study of Medicare patients who were hospitalized for heart failure due to LVSD between 30 June 1995 and 30 September 1996. Follow-up for each patient began on the date of discharge of the index hospitalization and continued for 21 months.

Statistical analysis

The dependent variable (outcome) was the number of readmissions measured during seven distinct time periods (0–3 months, >3–6 months, >6–9 months, >9–12 months, >12–15 months, >15–18 months, and >18–21 months after discharge from the index hospitalization). The primary exposure variable was treatment group at time of index hospital discharge (ACEI at target dose [reference], ACEI at suboptimal dose, and no treatment with ACEI).

Frequency of readmission. Table 3 shows the rate ratios for the 3-month time period and for the entire follow-up period. There were 1119 readmissions (mean 1.8 per patient) for the entire 21-month follow up period. The mean number of readmissions per patient was 1.53 (177/116) for patients who were prescribed target doses of ACEI at the index discharge. The mean was 1.83 (673/367) for patients prescribed ACEI at lower than target dose, and 2.10 (269/128) for patients who were not prescribed ACE inhibitors. Patients not prescribed ACEI at discharge had a 70% increase in readmission rates [(RR) 1.70; 95% CI 1.40–2.07], and patients at less than target dose a 31% increase (RR 1.31; 95% CI 1.11–1.56) compared with patients prescribed the drug at target doses.

Lesson Summary

- When Y is a Poisson count, interested in estimating mean count, μ, or the rate, λ
- Interested in the relationship between regressors and the log mean or log rate

Descriptive Statistics

Continuous

Center/Spread
Histogram

Categorical

Frequency Table
Bar Chart

Regression Modeling

Model of the Count

$$\log(\mu) = \alpha + \beta\, x$$

Model of the Rate

$$\log(\lambda) = \log\left(\frac{\mu}{t}\right) = \alpha + \beta\, x$$

Poisson Regression Model of the Count

$$\log(\hat{\mu}) = a + b\, x$$

Poisson Regression Model of the Rate

$$\log(\hat{\mu}) = a + b\, x + \log(t)$$