Chapter 8
Generalized Linear Models

A new program for depression is instituted in the hopes of reducing the number of visits each patient makes to the emergency room in the year following treatment. Predictors include (among many others) treatment (yes/no), race, and drug and alcohol usage indices. A common and minimally invasive treatment for jaundice in newborns is exposure to light. Yet the cost of this is high, mainly because of longer hospital stays, which are expensive. Predictors of the cost include race, gestational age, and birthweight.

These analyses require special attention both because of the nature of the outcome variable (counts in the depression example and costs, which are positive and right-skewed, for the jaundice example) and because the models we would typically employ are not as straightforward as the linear models of Chap. 4.

On the other hand, many features of constructing an analysis are the same as we have seen previously. We have a mixture of categorical (treatment, race) and continuous predictors (drug usage, alcohol usage, gestational age, birthweight). There are the same issues of determining the goals of inference (prediction, risk estimation, and testing of specific parameters) and winnowing down of predictors to arrive at a final model as discussed in Chap. 10. And we can use tests and CIs in ways that are quite similar to those for previously described analyses.

We begin this chapter by discussing the two examples in a bit more detail and conclude with a look at how those examples, as well as a number of earlier ones, can be subsumed under the broader rubric of *generalized linear models*.

8.1 Example: Treatment for Depression

A new case-management program for depression is instituted in a local hospital that often has to care for the poor and homeless. A characteristic of this population is that they often access the health care system by arriving in the emergency room—an

E. Vittinghoff et al., *Regression Methods in Biostatistics*, Statistics for Biology and Health, DOI 10.1007/978-1-4614-1353-0_8,

expensive and overburdened avenue to receive treatment. Can the new treatment reduce the number of needed visits to the emergency room as compared to standard care? The recorded outcome variable is the number of emergency room visits for each patient in the year following treatment.

The primary goal of the analysis is to assess the treatment program, but emergency room usage varies greatly according to other factors. Secondary goals included association of emergency room usage with drug or alcohol abuse and to assess racial differences in use.

8.1.1 Statistical Issues

From a statistical perspective, we need to be concerned with the nature of the outcome variable: in the data set that motivated this example, about one-third of the observations are 0 (did not return to the emergency room within the year) and over half are either 0 or 1. This is highly nonnormal and cannot be transformed to be approximately normal—any transformation by an increasing function will merely move the one-third of the observations that are exactly 0 to another numerical value, but there will still be a "lump" of observations at that point consisting of one-third of the data. For example, a commonly recommended transformation for count data with zeros is $\log(y + 1)$. This transformation leaves the data equal to 0 unchanged since $\log(0 + 1) = 0$ and moves the observations at 1 to $\log(1 + 1) = \log(2)$, not appreciably reducing the nonnormality of the data. Over half the data take on the two values 0 and $\log(2)$.

Even if we can handle the nonnormal distribution, a typical linear model (as in Chap. 4) for the mean number of emergency room visits will be untenable. The mean number of visits must be a positive number and a linear model, especially with continuous predictors, may, for extreme values of the covariates, predict negative values. This is the same problem we encountered with models for the probability of an event in Sect. 5.1.

Another bothersome aspect of the analysis is that this is a hard-to-follow, transient population in generally poor health. It is not at all unusual to have subjects die or be unable to be contacted for obtaining follow-up information. So some subjects are only under observation (and hence eligible for showing up for emergency room visits) for part of the year.

Since not all the subjects are followed for the same periods of time, it is natural to think of a multiplicative model. In other words, if all else is equal, a subject that is followed for twice as long as another subject will have, on average, twice the emergency room utilization. This consideration, as well as the desire to keep the mean response positive, leads us to consider a model for the log of the mean response. Note that this is different from the mean of the log-transformed responses (See Problem 8.1, also Sects. 4.7.2 and 4.7.5).

8.1.2 Model for the Mean Response

To begin to write down the model more carefully, define Y_i as the number of emergency room visits for patient i and let $\mathrm{E}[Y_i]$ represent the average number of visits for a year. For the moment we will ignore the fact that the observation periods are unequal. The model we are suggesting is

$$\log \mathrm{E}[Y_i] = \beta_0 + \beta_1 \mathrm{RACE}_i + \beta_2 \mathrm{TRT}_i + \beta_3 \mathrm{ALCH}_i + \beta_4 \mathrm{DRUG}_i, \tag{8.1}$$

or equivalently (using an exponential, i.e., anti-log)

$$\mathrm{E}[Y_i] = \exp\{\beta_0 + \beta_1 \mathrm{RACE}_i + \beta_2 \mathrm{TRT}_i + \beta_3 \mathrm{ALCH}_i + \beta_4 \mathrm{DRUG}_i\}, \tag{8.2}$$

where β_0 is an intercept, RACE_i is 1 for nonwhites and 0 for whites, TRT_i is 1 for those in the treatment group and 0 for usual care, ALCH_i is a numerical measure of alcohol usage and DRUG_i is a numerical measure of drug usage. We are primarily interested in β_2, the treatment effect.

Since the mean value is not likely to be exactly zero (otherwise, there is nothing to model), using the log function is mathematically acceptable (as opposed to trying to log transform the original counts, many of which are zero). Also, we can now reasonably hypothesize models like (8.1) that are linear (for the log of the mean) in ALCH_i and DRUG_i since the exponential in (8.2) keeps the mean value positive.

This is a model for the number of emergency room visits per year. What if the subject is only followed for half a year? We would expect their counts to be, on average, only half as large. A simple way around this problem is to model the mean count per unit time instead of the mean count, irrespective of the observation time. Let t_i denote the observation time for the ith patient. Then, the mean count per unit time is $\mathrm{E}[Y_i]/t_i$ and (8.1) can be modified to be

$$\log\left(\mathrm{E}[Y_i]/t_i\right) = \beta_0 + \beta_1 \mathrm{RACE}_i + \beta_2 \mathrm{TRT}_i + \beta_3 \mathrm{ALCH}_i + \beta_4 \mathrm{DRUG}_i, \tag{8.3}$$

or equivalently (using the fact that $\log[Y/t] = \log Y - \log t$)

$$\log \mathrm{E}[Y_i] = \beta_0 + \beta_1 \mathrm{RACE}_i + \beta_2 \mathrm{TRT}_i + \beta_3 \mathrm{ALCH}_i + \beta_4 \mathrm{DRUG}_i + \log t_i. \tag{8.4}$$

The term $\log t_i$ on the right-hand side of (8.4) looks like another covariate term, but with an important exception: there is no coefficient to estimate analogous to the β_3 or β_4 for the alcohol and drug covariates. Thinking computationally, if we used it as a predictor in a regression-type model, a statistical program like Stata would automatically estimate a coefficient for it. But, by construction, we know it must enter the equation for the mean with a coefficient of exactly 1. For this reason, it is called an *offset* instead of a covariate and when we use a package like Stata, it is designated as an offset and not a predictor.

8.1.3 Choice of Distribution

Lastly, we turn to the nonnormality of the distribution. Typically, we describe count data using the Poisson distribution. Directly modeling the data with a distribution appropriate for counts recognizes the problems with discreteness of the outcomes (e.g, the "lump" of zeros). While the Poisson distribution is hardly ever ultimately the correct distribution to use in practice, it gives us a place to start.

We are now ready to specify a model for the data, accommodating the three issues: nonnormality of the data, mean required to be positive, and unequal observation times. We start with the distribution of the data. Let λ_i denote the mean rate of emergency room visits per unit time, so that the mean number of visits for the ith patient is given by $\lambda_i t_i$. We then assume that Y_i has a Poisson distribution with log of the mean given by

$$\begin{aligned}\log \mathrm{E}[Y_i] &= \log[\lambda_i t_i] \\ &= \log \lambda_i + \log t_i \\ &= \beta_0 + \beta_1 \mathrm{RACE}_i + \beta_2 \mathrm{TRT}_i + \beta_3 \mathrm{ALCH}_i + \beta_4 \mathrm{DRUG}_i + \log t_i. \quad (8.5)\end{aligned}$$

This shows us that the main part of the model (consisting of all the terms except for the offset $\log t_i$) is modeling the rate of emergency room visits per unit time:

$$\log[\lambda_i] = \beta_0 + \beta_1 \mathrm{RACE}_i + \beta_2 \mathrm{TRT}_i + \beta_3 \mathrm{ALCH}_i + \beta_4 \mathrm{DRUG}_i, \quad (8.6)$$

or, exponentiating both sides,

$$\lambda_i = \exp\{\beta_0 + \beta_1 \mathrm{RACE}_i + \beta_2 \mathrm{TRT}_i + \beta_3 \mathrm{ALCH}_i + \beta_4 \mathrm{DRUG}_i\}. \quad (8.7)$$

8.1.4 Interpreting the Parameters

The model in (8.7) is a multiplicative one, as we saw for the Cox model in Chap. 6, and has a similar style of interpretation. Recall that RACE_i is 1 for nonwhites and 0 for whites and suppose the race coefficient is estimated to be $\hat{\beta}_1 = -0.5$. The mean rate per unit time for a white person divided by that of a nonwhite (assuming treatment group, and alcohol and drug usage indices are all the same) would be

$$\begin{aligned}&\frac{\exp\{\beta_0 + 0 + \beta_2 \mathrm{TRT} + \beta_3 \mathrm{ALCH} + \beta_4 \mathrm{DRUG}\}}{\exp\{\beta_0 - 0.5 + \beta_2 \mathrm{TRT} + \beta_3 \mathrm{ALCH} + \beta_4 \mathrm{DRUG}\}} \\ &= \frac{e^{\beta_0} e^{0} e^{\beta_2 \mathrm{TRT}} e^{\beta_3 \mathrm{ALCH}} e^{\beta_4 \mathrm{DRUG}}}{e^{\beta_0} e^{-0.5} e^{\beta_2 \mathrm{TRT}} e^{\beta_3 \mathrm{ALCH}} e^{\beta_4 \mathrm{DRUG}}}\end{aligned}$$

$$= \frac{e^0}{e^{-0.5}}$$
$$= e^{0.5} \approx 1.65. \tag{8.8}$$

So the interpretation is that, after adjustment for treatment group and alcohol and drug usage, whites tend to use the emergency room at a rate 1.65 that of the nonwhites. Said another way, the average rate of usage for whites is 65% higher than that for non-whites. Similar, multiplicative, interpretations apply to the other coefficients.

In summary, to interpret the coefficients when modeling the log of the mean, we need to exponentiate them and interpret them in a multiplicative or ratio fashion. In fact, it is often good to think ahead to the desired type of interpretation. Proportional increases in the mean response due to covariate effects are sometimes the most natural interpretation and are easily incorporated by planning to use such a model.

8.1.5 Further Notes

Models like the one developed in this section are often called Poisson regression models, named after the distribution assumed for the counts. A feature of the Poisson distribution is that the mean and variance are required to be the same. So, if the mean number of emergency room visits per year is 1.5, for subjects with a particular pattern of covariates, then the variance would also be 1.5 and the standard deviation would be the square root of that or about 1.23 visits per year. Ironically, the Poisson distribution often fails to hold in practice since the variability in the data often exceeds that of the mean. A common solution (where appropriate) is to assume that the variance is proportional to the mean, not exactly equal to it, and estimate the proportionality factor, which is called the *scale parameter*, from the data. For example, a scale parameter of 2.5 would mean that the variance was 2.5 times larger than the mean and this fact would be used in calculating standard errors, hypothesis tests, and confidence intervals. When the scale parameter is greater than 1, meaning that the variance is larger than that assumed by the named distribution, the data are termed *overdispersed*. Another solution is to choose a different distribution. For example, the Stata package has a negative binomial (a different count data distribution) regression routine, in which the variance is modeled as a quadratic function of the mean.

The use of log time as an offset in model (8.5) may seem awkward. Why not just divide each count by the observation period and analyze Y_i / t_i? The answer is that it makes it harder to think about and specify the proper distribution. Instead of having count data, for which there are a number of statistical distributions to choose from, we would have a strange, hybrid distribution, with "fractional" counts, e.g., with an

observation period of 0.8 of a year, we would could obtain values of 0, 1.25 (which is 1 divided by 0.8), 2.5, 3.75, etc. With a different observation period, a different set of values would be possible.

8.2 Example: Costs of Phototherapy

About 60% of newborns become jaundiced, i.e., the skin and whites of the eyes turn yellow in the first few days after birth. Newborns become jaundiced because they have an increase in bilirubin production due to increased red blood cell turnover and because it takes a few days for their liver (which helps eliminate bilirubin) to mature. Newborns are treated for jaundice because of the possibility of bilirubin-induced neurological damage. What are the costs associated with this treatment and are costs also associated with race, the gestational age of the baby, and the birthweight of the baby?

Our outcome will be the total cost of health care for the baby during its first month of life. Cost is a positive variable and is almost invariably highly skewed to the right. A common remedy is to log transform the costs and then fit a multiple regression model. This is often highly successful as log costs are often well-behaved statistically, i.e., approximately normally distributed and homoscedastic. This is adequate if the main goal is to test whether one or more risk factors are related to cost.

However, if the goal is to understand the determinants of the actual *cost* of health care, then it is only the mean cost that is of interest (since mean cost times the number of newborns is the total cost to the health care system). One strategy is to perform the analysis on the log scale and then back transform (using an exponential) to get things back on the original cost scale.

However, since the log of the mean is not the same as the mean of the log, back-transforming an analysis on the log scale does not directly give results interpretable in terms of mean costs. Instead they are interpretable as models for median cost (Goldberger 1968). The reasoning behind this is as follows. If the log costs are approximately normally distributed, then the mean and median are the same. Since monotonic transformations preserve medians (the log of the median value *is* the median of the log values) back-transforming using exponentials gives a model for median cost. There are methods for getting estimates of the mean via adjustments to the back transformation (Bradu and Mundlak 1970) but there are also alternatives.

One alternative is to adopt the approach of the previous section: model the mean and assume a reasonable distribution for the data. What choices would we need to make for this situation?

A reasonable starting point is to observe that the mean cost must be positive. Additive and linear models for positive quantities can cause the problem of negative predicted values and hence multiplicative models incorporating proportional changes are commonly used. For cost, this is often a more natural characterization, i.e., “low birthweight babies cost 50% more than normal birthweight babies”

and is likely to be more stable than modeling absolute changes in cost (locations with very different costs of care are unlikely to have the same differences in costs, but may have the same ratio of costs). As in the previous section, that would lead to a model for the log of the mean cost (similar to but not the same as log-transforming cost).

8.2.1 Model for the Mean Response

More precisely, let us define Y_i as the cost of health care for infant i during its first month and let $\mathrm{E}[Y_i]$ represent the average cost. Our model would then be

$$\log \mathrm{E}[Y_i] = \beta_0 + \beta_1 \mathrm{RACE}_i + \beta_2 \mathrm{TRT}_i + \beta_3 \mathrm{GA}_i + \beta_4 \mathrm{BW}_i, \tag{8.9}$$

or equivalently (using an exponential)

$$\mathrm{E}[Y_i] = \exp\{\beta_0 + \beta_1 \mathrm{RACE}_i + \beta_2 \mathrm{TRT}_i + \beta_3 \mathrm{GA}_i + \beta_4 \mathrm{BW}_i\}, \tag{8.10}$$

where β_0 is an intercept, RACE_i is 0 for whites and 1 for non-whites, TRT_i is 1 for those receiving phototherapy and 0 for those who do not, GA_i is the gestational age of the baby, and BW_i is its birthweight. We are primarily interested in β_2, the phototherapy effect.

8.2.2 Choice of Distribution

The model for the mean for the jaundice example is virtually identical to that for the depression example in Sect. 8.1.2. But the distributions need to be different since cost is a continuous variable, while number of emergency room visits is discrete. There is no easy way to know what distribution might be a good approximation for such a situation, without having the data in hand. However, it is often the case that the standard deviation in the data increases proportionally with the mean. This situation can be diagnosed by looking at residual plots (as described in Chap. 4) or by plotting the standard deviations calculated within subgroups of the data versus the means for those subgroups. In such a case, a reasonable choice is the gamma distribution, which is a flexible distribution for positive, continuous variables that incorporates the assumption that the standard deviation is proportional to the mean.

When we are willing to use a gamma distribution as a good approximation to the distribution of the data, we can complete the specification of the model as follows. We assume that Y_i has a gamma distribution with mean, $\mathrm{E}[Y_i]$, given by

$$\log \mathrm{E}[Y_i] = \beta_0 + \beta_1 \mathrm{RACE}_i + \beta_2 \mathrm{TRT}_i + \beta_3 \mathrm{GA}_i + \beta_4 \mathrm{BW}_i. \tag{8.11}$$

8.2.3 *Interpreting the Parameters*

Since the model is a model for the log of the mean, the parameters have the same interpretation as in the previous section. For example, if $\hat{\beta}_2 = 0.5$ (positive since phototherapy increases costs) then the interpretation would be that, adjusting for race, gestational age, and birthweight, the cost associated with babies receiving phototherapy was $\exp(0.5) \approx 1.65$ as high as those not receiving it.

8.3 Generalized Linear Models

The examples in Sects. 8.1 and 8.2 have been constructed to emphasize the similarity of the models (compare Subsects. 8.1.4 and 8.2.3) for two very different situations. So even with very different distributions (Poisson versus gamma) and different statistical analyses, they have much in common.

A number of statistical packages, including Stata, have what are called *generalized linear model* commands that are capable of fitting linear, logistic, Poisson regression and other models. The basic idea is to let the data analyst tailor the analysis to the data rather than having to transform or otherwise manipulate the data to fit an analysis. This has significant advantages in situations like the phototherapy cost example where we want to model the outcome without transformation.

Fitting a GLM involves making a number of decisions:

(1) What is the distribution of the data (for a fixed pattern of covariates)?
(2) What function will be used to *link* the mean of the data to the predictors?
(3) Which predictors should be included in the model?

In the examples in the preceding sections we used Poisson and gamma distributions, we used a log function of the mean to give us a linear model in the predictors and our choice of predictors was motivated by the subject matter. Note that choices on the predictor side of the equation are largely independent of the first two choices.

In previous chapters, we have covered linear and logistic regression. In linear regression, we modeled the mean directly and assumed a normal distribution. This is using an *identity link function*, i.e., we modeled the mean identically, without transforming it. In logistic regression, we modeled the log of the odds, i.e., $\log(p/[1-p])$, and assumed a binomial or binary outcome. If the outcome is coded as zero for failure and one for success, then the average of the zeros and ones is p, the probability of success. In that case, we used a *logit link* to link the mean, p, to the predictors.

Generalized linear model commands give large degrees of flexibility in the choice of each of the features of the model. For example, current capabilities in Stata are to handle six distributions (normal, binomial, Poisson, gamma, negative binomial,

Table 8.1 Count regression example assuming a Poisson distribution

```
glm shared_syr i.homeless, family(poisson) link(log) eform

Generalized linear models                         No. of obs      =        121
Optimization     : ML                             Residual df     =        119
                                                  Scale parameter =          1
Deviance         =   1511.02467                   (1/df) Deviance =  12.69769
Pearson          =  3586.309617                   (1/df) Pearson  =  30.13706

Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = ln(u)                   [Log]

                                                  AIC             =  13.33909
Log likelihood   = -805.0147598                   BIC             =  940.3256

------------------------------------------------------------------------------
             |                 OIM
  shared_syr |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  1.homeless |   3.270615   .3985062    9.73   0.000     2.575819    4.152825
------------------------------------------------------------------------------
```

and inverse gaussian), and ten link functions (including identity, log, logit, probit, power functions).

8.3.1 Example: Risky Drug Use Behavior

Here is an example of modeling risky drug use behavior (sharing syringes) among drug users. The outcome is the number of times the drug user shared a syringe (`shared_syr`) in the past month (values ranged from 0 to 60!) and we will consider a single predictor, whether or not the drug user was homeless. Table 8.1 gives the results assuming a Poisson distribution. The Stata command, `glm`, specifies a Poisson distribution and a log link and we have specified the option `eform`, which automatically exponentiates the coefficients. The output contains a number of standard elements, including estimated coefficients, standard errors, Z-tests, P-values, and CIs. The homeless coefficient is highly statistically significant, with a value of about 3.27, meaning that being homeless is associated with over three times more use of shared syringes than nonhomeless.

However, these data are highly variable and the Poisson assumption of equal mean and variance is dubious. If we specify the `vce(robust)` a robust variance estimate will be used in the calculation of the standard errors. Just as described in Chap. 7, the robust variance estimate gives valid standard errors even when the assumed form of the variance is incorrect, in this case that the variance is equal to the mean.

Table 8.2 gives the result with the robust standard errors, which is not quite statistically significant. Standard errors have increased approximately fivefold using the `vce(robust)` option, so the assumption of a Poisson distribution is far from correct. In the terminology of generalized linear models, these data are highly

Table 8.2 Count regression example with scaled standard errors

```
. glm shared_syr i.homeless, family(poisson) link(log) eform vce(robust)

Generalized linear models                         No. of obs      =        121
Optimization     : ML                             Residual df     =        119
                                                  Scale parameter =          1
Deviance         =   1511.02467                   (1/df) Deviance =  12.69769
Pearson          =  3586.309617                   (1/df) Pearson  =  30.13706

Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = ln(u)                   [Log]

                                                  AIC             =  13.33909
Log pseudolikelihood = -805.0147598               BIC             =  940.3256

------------------------------------------------------------------------------
             |               Robust
  shared_syr |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  1.homeless |   3.270615   2.005987    1.93   0.053     .9830072    10.88184
------------------------------------------------------------------------------
```

overdispersed, because the variance is much larger than that assumed for a Poisson distribution.

This example serves as a warning not to make strong assumptions, such as those embodied in using a Poisson distribution, blindly. It is wise at least to make a sensitivity check by using the robust variance estimator for count data as well as for binomial data with denominators other than 1 (with binary data, with a denominator of 1, no overdispersion is possible). Also, when there are just a few covariate patterns and subjects can be grouped according to their covariate values, it is wise to plot the variance within such groups versus the mean within the group to display the variance to mean relationship graphically. The mean values for the `shared_syr` variable are 4.7 and 1.4 for the homeless and nonhomeless groups, respectively, with corresponding standard deviations of 13.7 and 5.5. So the standard deviations are roughly three times the mean, as reflected by the robust standard errors being much larger in Table 8.2 compared to Table 8.1.

An alternative distribution for count data that allows more variability than the Poisson is the negative binomial distribution. Table 8.3 shows the negative binomial distribution fit. The estimated effect of being homeless is the same as the Poisson fit and the standard errors, *p*-value and CI are all similar to those in Table 8.2, which uses a robust standard error.

8.3.2 Modeling Data with Many Zeros

When analyzing count *or* numerical outcome data, it is not unusual to discover a large percentage of the data being zero. For example, in a study following the members of a health plan for use of emergency room visits, the vast majority would be zero, with the nonzero outcomes taking on integer values. If we change the

Table 8.3 Count regression using a negative binomial distribution

```
. glm  shared_syr i.homeless, family(nbinomial ml) ef

Generalized linear models                         No. of obs      =        121
Optimization     : ML                             Residual df     =        119
                                                  Scale parameter =          1
Deviance         =  58.10151246                   (1/df) Deviance =   .488248
Pearson          =  94.44640018                   (1/df) Pearson  =  .7936672

Variance function: V(u) = u+(14.1206)u^2          [Neg. Binomial]
Link function    : g(u) = ln(u)                   [Log]

                                                  AIC             =  2.591876
Log likelihood   = -154.8084869                   BIC             = -512.5976

------------------------------------------------------------------------------
             |                 OIM
  shared_syr |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  1.homeless |   3.270616   2.270502    1.71   0.088      .8389088     12.751
------------------------------------------------------------------------------
Note: Negative binomial parameter estimated via ML and treated as fixed once
estimated.
```

outcome to hospitalization costs, again the vast majority would be zero, but the nonzero values would likely be positive and skewed right. For the syringe sharing data, 78% of the outcomes are zero. How can these be handled in practice? As noted in Sect. 8.1.1, a transformation of the outcome will not help.

A simple strategy is to build separate models for the zeros and the nonzero values. These are sometimes called conditional, two-part or "hurdle" models, the latter name arising because, after the outcomes "hurdle" the value of zero, a different model is used. For the analysis of hospitalization costs, we could use a logistic regression for the probability of the cost being zero. And for the nonzero costs, we could fit a GLM assuming a gamma distribution or we could log transform the outcome to try to make it approximately normally distributed. The same predictors can be in both models or we can model each outcome with its own collection of predictors.

For the syringe sharing data, we could use a logistic regression to model the chance of sharing zero syringes with the predictor of being homeless. But what model could we use for the nonzero data? It does not fit any usual count data model, because there are no zeros allowed. Fortunately, Stata can accommodate either a Poisson or negative binomial distribution which has been *truncated* to only allow nonzero values through its `ztp` (zero-truncated Poisson) or `ztnb` (zero-truncated negative binomial) regression commands.

Table 8.4 shows the two fits, first modeling the probability of no syringe sharing with the predictor of being homeless and a logistic regression and then the number of times syringes are shared, using a zero-truncated negative binomial distribution. Because the fits provide two tests of the homeless effect based on the same data, we could use a Bonferroni correction (see Sect. 4.3.4) and test each at a significance level of 0.025 instead of 0.05; correspondingly, we have used the `level` option

Table 8.4 Fitting a two-part model to the syringe sharing data

```
. gen share0=(shared_syr==0)
. logistic share0 i.homeless, level(97.5)

Logistic regression                               Number of obs   =        124
                                                  LR chi2(1)      =       3.19
                                                  Prob > chi2     =     0.0740
Log likelihood = -67.012808                       Pseudo R2       =     0.0233

------------------------------------------------------------------------------
      share0 | Odds Ratio   Std. Err.      z    P>|z|  [97.5% Conf. Interval]
-------------+----------------------------------------------------------------
  1.homeless |   .4676114   .2019848   -1.76   0.078    .1775875    1.231283
------------------------------------------------------------------------------

. ztnb shared_syr i.homeless if shared_syr>0, irr level(97.5)

Zero-truncated negative binomial regression     Number of obs   =         27
                                                LR chi2(1)      =       1.02
Dispersion     = mean                           Prob > chi2     =     0.3133
Log likelihood = -91.29219                      Pseudo R2       =     0.0055

------------------------------------------------------------------------------
  shared_syr |        IRR   Std. Err.      z    P>|z|  [97.5% Conf. Interval]
-------------+----------------------------------------------------------------
  1.homeless |   2.122108   1.501492    1.06   0.288    .4345308    10.36369
-------------+----------------------------------------------------------------
    /lnalpha |   1.584369   1.081463                     -.8396253    4.008363
-------------+----------------------------------------------------------------
       alpha |   4.876212   5.273443                      .4318723    55.05666
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01)= 413.32 Prob>=chibar2 = 0.000
```

to set the CIs to have 97.5% confidence. The logistic model estimates the odds ratio of *not sharing* a needle. Perhaps easier to interpret, the homeless have odds of sharing a needle which are a little over two times higher than the nonhomeless ($2.1385 = 1/.4676$). The zero-truncated regression estimate indicates that, among those that do share needles, the homeless share syringes at a rate a little over two times more often than the nonhomeless. Because of the Bonferroni correction, each of the tests would require a p-value of 0.025 to be declared statistically significant and neither is.

These two-part or zero-inflated (below) modeling approaches are especially attractive in situations where different predictors might influence the two parts of the model. For example, what determines whether or not someone is willing to share needles may be quite different from what determines how frequently they share needles when they do.

Another approach to modeling count data with many zeros is to use what are called *zero-inflated* models. Rather than breaking the data into two parts, a zero-inflated approach uses an underlying model in which two processes are operating: first a process that generates the zeros (like the logistic regression above) and then a count data model, such as the Poisson. This is slightly different from the two-part model since a zero in the data could have arisen either from the zero-generation process or from the count data process, which just happens to generate

Table 8.5 Fitting a zero-inflated negative binomial model to the syringe sharing data

```
. zinb shared_syr i.homeless, inflate(i.homeless) irr level(97.5)

Zero-inflated negative binomial regression      Number of obs   =         121
                                                Nonzero obs     =          27
                                                Zero obs        =          94

Inflation model = logit                         LR chi2(1)      =        1.02
Log likelihood  =  -154.112                     Prob > chi2     =      0.3133

------------------------------------------------------------------------------
  shared_syr |        IRR   Std. Err.      z    P>|z|   [97.5% Conf. Interval]
-------------+----------------------------------------------------------------
shared_syr   |
  1.homeless |   2.122107   1.501493    1.06   0.288    .4345298     10.3637
-------------+----------------------------------------------------------------
inflate      |
  1.homeless |  -.7708614   .7434234   -1.04   0.300   -2.437173     .8954498
       _cons |   .6342794   1.161001    0.55   0.585   -1.967991     3.236549
-------------+----------------------------------------------------------------
    /lnalpha |   1.584377    1.08147    1.47   0.143   -.8396342     4.008387
-------------+----------------------------------------------------------------
       alpha |    4.87625    5.27352                     .4318685     55.05801
------------------------------------------------------------------------------
```

a zero. These models are more natural for some situations. For example, consider modeling the number of open nurse anesthetist positions per hospital, similar to the study of Merwin et al. (2009), with predictors being the log of the number of surgeries, log of the average daily number of patients, log of the number of operating rooms and the state in which it is located. The number of open positions could be zero because the hospital does not hire nurse anesthetists or because they do, but they have no open positions. Zero-inflated models can be used in situations in which we can justify the underlying two processes or in situations in which we merely need to accommodate the large percentage of zero outcome values.

Table 8.5 shows the results from fitting a zero-inflated negative binomial model, where the `inflate` option gives the predictors for the underlying process that generates the zeros and again we have set the level to 97.5% to accommodate the two tests. The estimates and interpretations are very similar to the two-part fit above, with the effect of being homeless on the count data model being identical and the effect of being homeless on the zero model being very similar. Table 8.5 reports the log odds of not sharing a syringe as -0.7708, which corresponds to an odds ratio of $\exp\{-0.7708\} = 0.4626$, similar to Table 8.4.

8.3.3 Example: A Randomized Trial to Reduce Risk of Fracture

Osteoporosis (roughly porous bone, from the Greek) is a condition in which bones become weak and brittle and readily susceptible to fracture. It primarily affects postmenopausal women and can lead to chronic pain, skeletal deformities, and

Table 8.6 Fracture risk by fall risk and treatment group

```
glm numnosp ibn.trt_fall, family(poisson) offset(logyears) vce(robust) noconstant ef

Generalized linear models                         No. of obs      =      6369
Optimization     : ML                             Residual df     =      6365
                                                  Scale parameter =         1
Deviance         =  4116.885884                   (1/df) Deviance =  .6468006
Pearson          =  8002.406864                   (1/df) Pearson  =  1.257252

Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = ln(u)                   [Log]

                                                  AIC             =  .9180298
Log pseudolikelihood = -2919.465795               BIC             = -51635.41

------------------------------------------------------------------------------
             |               Robust
     numnosp |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    trt_fall |
          1  |    .041815   .0022419  -59.21   0.000     .0376439    .0464482
          2  |   .0340865   .0020225  -56.95   0.000     .0303444    .0382902
          3  |   .0509974   .0056819  -26.71   0.000     .0409931    .0634431
          4  |   .0521462   .0055934  -27.54   0.000      .042259    .0643466
    logyears |   (offset)
------------------------------------------------------------------------------
```

increased risk of death. The Fracture Intervention Trial (Black et al. 1996a) was a randomized controlled trial among postmenopausal women that showed that alendronate (a drug that increases bone density) was able to reduce the risk of fracture.

Falling is a major cause of fractures, but would alendronate prevent fractures from an event as traumatic as a fall? To answer this, women at high risk of falling were identified by poor performance on the "Timed Up and Go" test, which measures how long it takes to stand up from an armchair, walk 3 m, return and sit down, and has been shown to be a predictor of the risk of falling. The effect of alendronate on the number of nonspine fractures (`numnosp`) was then estimated separately for the high and low risk of falling groups.

Women were not followed for the same amount of time, so we use an offset of log of years in the trial (`logyears`). To get estimated rates for each of the groups, we created a four level, categorical variable (`trt_fall`) with values 1–4 representing, respectively, the low risk/placebo, low risk/alendronate, high risk/placebo, and high risk/alendronate groups. We use the `ibn.` prefix for the `trt_fall` variable (so there is no omitted baseline reference group) and `noconstant` options to force Stata to include all four groups and to not fit a constant term.

We fit the model using the `glm` command, specifying a Poisson distribution, using robust standard errors (in case of overdispersion), and reporting the results with the `eform` option to directly display the yearly fracture rates. The output is displayed in Table 8.6.

In the low risk of falling groups, the yearly rates of fracture are much less (0.042 for the placebo group and 0.034 for the alendronate groups). These correspond to about 4.2 and 3.4 fractures per 100 women over a year. In the high risk of falling groups, the rates are higher and about the same as one another (about 5.1 and 5.2 fractures per 100 woman years). We wish to compare the risk difference between the treated and untreated groups over the average followup time of 3.8 years. This is of interest because, unlike a relative risk or odds ratio, it is easily related to the number of fractures prevented by treatment.

To make the formal comparison, we fit a model with effects for treatment (`trt`), fall risk category (`fall_risk`), and their interaction. The `margins` command can be conveniently used with the `@` operator to compare the treatment groups within the fall risk categories. The results are given in Table 8.7.

The results show that, over a period of 3.8 years, the risk associated with being treated by alendronate in the low fall risk group is about 0.029 less than the untreated group. In other words, in the low fall risk group, treatment with alendronate prevents about 2.9 fractures per 100 women over a 3.8-year treatment period. However, in the high risk group, the difference between alendronate and the placebo group is not statistically significant and has a small estimated effect (the drug is estimated to increase the number of fractures per 100 women over 3.8 years by about 0.4 fractures). Because the high risk group is smaller, the confidence interval is wide, and so we cannot rule out clinically important differences.

The analysis above is appropriate if the focus is, a priori, on estimating the treatment effects separately in the low and high risk of falling groups. In this case, the results also suggest that there is a difference in the treatment effect in the two groups. However, if the goal is to directly compare the treatment effects in the low and high risk groups, a better approach is to test for the interaction between risk of falling and treatment (see the third point in Sect. 4.6). The test of interaction with this data in Table 8.7 does not provide strong evidence for a difference in treatment effects, with a p-value of 0.19. This is a caution not to interpret statistical significance of an effect in one group and lack of statistical significance in another group as evidence for a difference in the effects in the two groups. This is especially true with unequal sample sizes, such as in this example.

8.3.4 Relationship of Mean to Variance

The key to use of a GLM program is the specification of the relationship of the mean to the variance. This is the main information used by the program to fit a model to data when a distribution is specified. As noted above, this relationship can often be assessed by residual plots or plots of subgroup standard deviations versus means. Table 8.8 gives the assumed variance to mean relationship, distributional name, and situations in which the common choices available in Stata would be used.

Table 8.7 Fracture risk treatment comparisons within fall risk categories

```
glm numnosp i.trt##i.fall_risk, family(poisson) offset(logyears)
vce(robust) ef


Generalized linear models                         No. of obs      =       6369
Optimization     : ML                             Residual df     =       6365
                                                  Scale parameter =          1
Deviance         =  4116.885884                   (1/df) Deviance =  .6468006
Pearson          =  8002.406864                   (1/df) Pearson  =  1.257252

Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = ln(u)                   [Log]

                                                  AIC             =  .9180298
Log pseudolikelihood = -2919.465795               BIC             = -51635.41
------------------------------------------------------------------------------
             |               Robust
    numnosp |        IRR   Std. Err.      z    P>|z|   [95% Conf. Interval]
-------------+----------------------------------------------------------------
        1.trt |   .8151755   .0651883   -2.56   0.011   .6969183    .9534993
  1.fall_risk |   1.219596   .1507959    1.61   0.108   .9571279     1.55404
              |
trt#fall_risk |
        1 1   |   1.254364   .2183953    1.30   0.193   .8917067    1.764513
              |
        _cons |    .041815   .0022419  -59.21   0.000   .0376439    .0464482
     logyears |          1  (offset)
------------------------------------------------------------------------------

. margins r.trt@fall_risk

Contrasts of adjusted predictions
Model VCE    : Robust

Expression   : Predicted mean numnosp, predict()
------------------------------------------------
             |         df        chi2     P>chi2
-------------+----------------------------------
trt@fall_risk |
  (1 vs 0) 0  |          1        6.55     0.0105
  (1 vs 0) 1  |          1        0.02     0.8854
       Joint  |          2        6.57     0.0374
------------------------------------------------

------------------------------------------------------------------
             |            Delta-method
             |   Contrast   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
trt@fall_risk |
  (1 vs 0) 0  |  -.0293294   .0114584     -.0517874   -.0068713
  (1 vs 0) 1  |   .0043597   .0302577     -.0549444    .0636637
------------------------------------------------------------------
```

8.3.5 Non-Linear Models

Not every model fits under the GLM umbrella. Use of the method depends on finding a transformation of the mean for which the predictors enter as a linear model, which may not always be possible. For example, in drug pharmacokinetics, a common model for the mean concentration of a drug in blood, Y, as a function of time, t, is:

Table 8.8 Common distributional choices for generalized linear models in Stata

Distribution	Variance to mean[a]	Sample situation
Normal	Constant σ^2	Linear regression
Binomial	$\sigma^2 = n\mu(1-\mu)$	Successes out of n trials
OD[b] Binomial	$\sigma^2 \propto n\mu(1-\mu)$	Clustered success data
Poisson	$\sigma^2 = \mu$	Count data, variance equals mean
OD Poisson	$\sigma^2 \propto \mu$	Count data, variance proportional to mean
Negative binomial	$\sigma^2 = \mu + \mu^2/k$	Count data, variance quadratic in the mean
Gamma	$\sigma \propto \mu$	Continuous data, standard deviation proportional to mean

[a]Mean is denoted by μ and the variance by σ^2.
[b]Over-dispersed.

$$\mathrm{E}[Y] = \mu_1 \exp\{-\lambda_1 t\} + \mu_2 \exp\{-\lambda_2 t\}. \tag{8.12}$$

In addition to time, we might have other predictors such as drug dosage or gender of the subject. However, there is no transformation that will form a linear predictor, even without the inclusion of dose and gender effects, and so a generalized linear model is not possible. As a consequence, more care must be taken in deciding how to incorporate the effects of predictor variables and building regression models is thus more complicated for nonlinear models. Software for fitting nonlinear models is relatively common for approximately normally distributed outcomes (such as `nl` in Stata) but less so for nonnormally distributed outcomes.

8.4 Sample Size for the Poisson Model

Section 5.7 provides formulas for calculating sample size, power, and minimum detectable effects for the logistic model. Similar results hold for the Poisson model. To compute the sample size that will provide power of γ in two-sided tests with type-1 error of α to reject the null hypothesis $\beta_j = 0$ for the effect of a predictor X_j, accounting for the loss of precision arising from multiple predictors, we can use

$$n = \frac{(z_{1-\alpha/2} + z_\gamma)^2 \theta}{(\beta_j^a \sigma_j)^2 \mu \left(1 - \rho_j^2\right)}, \tag{8.13}$$

where β_j^a is the hypothesized value of β_j under the alternative, $z_{1-\alpha/2}$ and z_γ are the quantiles of the standard normal distribution corresponding to the specified type-1 error and power, ρ_j is the multiple correlation of X_j with the other covariates, μ is the marginal mean of the count outcome, and θ is the scale parameter introduced in Sect. 8.3.1, defined as the ratio of variance of the outcome to μ. When X_j is binary with prevalence f_j, $\sigma_{x_j} = \sqrt{f_j(1-f_j)}$. For problems with predetermined n, power is given by

$$\gamma = 1 - \Phi\left[z_{1-\alpha/2} - \beta_j^a s_{x_j}\sqrt{n\mu\left(1-\rho_j^2\right)/\theta}\right]. \tag{8.14}$$

Finally, the minimum detectable effect (on the log-mean scale) is

$$\pm\,\beta_j^a = \frac{z_{1-\alpha/2} + z_\gamma}{s_{x_j}\sqrt{n\mu(1-\rho_j^2)/\theta}}. \tag{8.15}$$

Some additional points:

- Sample size (8.13) and minimum detectable effect (8.15) calculations simplify considerably when we specify $\alpha = 0.05$ and $\gamma = 0.8$, β_j^a is the effect of a one standard deviation increase in continuous x_j, and we do not need to penalize for covariate adjustment. However, we do assume that over-dispersion may still need to be taken into account, via the parameter θ. In that case,

$$n = \frac{7.849}{(\beta_j^a)^2\mu/\theta}. \tag{8.16}$$

  For the minimum detectable effect, we have

$$\pm\,\beta_j^a = \frac{2.802}{\sqrt{n\mu/\theta)}}. \tag{8.17}$$

  For 90% power, substitute 10.51 for 7.849 and 3.242 for 2.802.
- Similarly, for a two-arm clinical trial with equal allocation to arms, so that β_j^a is the log rate ratio for treatment, and $s_{x_j}^2 = 0.25$, we can calculate

$$n = \frac{4 \times 7.849}{(\beta_j^a)^2\mu/\theta}. \tag{8.18}$$

  For the minimum detectable effect, we have

$$\pm\,\beta_j^a = \frac{2 \times 2.802}{\sqrt{n\mu/\theta}}. \tag{8.19}$$

- Power calculations using (5.17) simplify analogously, but still require a statistical calculator or computer package to evaluate the normal cumulative distribution function $\Phi(\cdot)$.
- To our knowledge, these computations are not implemented in any statistical packages. However, (8.13)–(8.15) can easily be programmed in Stata, R, or Excel, or evaluated by hand if values of $z_{1-\alpha/2}$, z_γ, and $\Phi(\cdot)$ are available.
- As in calculations for other models, we need to use $|\beta_j^a|$ in (8.13) and (8.14) if $\beta_j^a < 0$.

Table 8.9 Sample size calculations for trial of behavioral intervention

```
. display (invnormal(.975) + invnormal(.9))^2 * 30 / ((log(0.5) * 0.5)^2 * 7.5)
349.91719
```

- The severe overdispersion evident in the example in Sect. 8.3.1 underlines the importance of obtaining a good estimate of θ, the scale parameter capturing overdispersion in (8.13)–(8.15). Note that $n \propto \theta$.
- The use of the variance inflation factor to account for covariate adjustment carries over to GLMs. However, there is no analog to the reduction in residual variance, so that the adjustment based on the variance inflation factor is less likely to be conservative for these models.
- $\mathrm{SE}(\hat{\beta}_j)$ is a large-sample approximation, and more exact small-sample computations using the t-distribution do not carry over from the linear model. Simulations of power may be a more reliable guide in those circumstances.
- Equations (8.13)–(8.15) are based on the assumption that the conditional mean of the outcome does not vary strongly across observations; methods based on more complicated calculations or simulation avoid this simplification and perform slightly better in some circumstances (Vittinghoff et al. 2009). However, errors from these sources are usually small compared to errors arising from uncertainty about the required inputs.
- The alternative calculations (4.22)–(4.24) presented in Sect. 4.8, which use an estimate $\tilde{\mathrm{SE}}(\hat{\beta}_j)$ based on published results for an appropriately adjusted model using $\tilde{n}$ observations, carry over directly. However, care must be taken to obtain the SE of the regression coefficient β_j, not the SE of the rate-ratio e^{β_j}. This can computed from a 95% CI for the rate-ratio as $\tilde{\mathrm{SE}}(\hat{\beta}_j) = \log(\mathrm{UL}/\mathrm{LL})/3.92$, where UL and LL are the upper and lower bounds. We must also ensure that β_j^a is based on the same predictor scale as in the published results.

To illustrate these calculations, suppose we are planning a randomized trial to assess the effectiveness of a behavioral intervention for reducing syringe sharing among drug users. Equal numbers will be allocated to the active intervention and a wait-list control, so that $f_j = 0.5$ and $s_{x_j} = \sqrt{0.5(1-0.5)} = 0.5$. Because the trial is randomized, we can assume that $\rho_j = 0$. Using the data shown in Tables 8.1 and 8.2, we estimate that among the wait-list controls, $\mu = 10$, and $\theta = 30$. We hypothesize that the intervention will reduce the frequency of sharing by 50%, so that overall, $\mu = 7.5$ and $\beta_j^a = \log 0.5$. In this case, we require power of 90% in a two-sided test with α of 5%.

Table 8.9 shows the sample size estimate of 350. This estimate has been inflated by a factor of $\theta = 30$ to account for overdispersion of the outcome. Clearly a naïve estimate assuming equality of the mean and variance would result in a badly underpowered trial.

8.5 Summary

The purpose of this chapter has been to outline the topic of GLMs, a class of models capable of handling a wide variety of analysis situations. Specification of the generalized linear model involves making three choices:

(1) What is the distribution of the data (for a fixed pattern of covariates)? This must be specified at least up to the variance to mean relationship.
(2) What function will be used to link the mean of the data to the predictors?
(3) Which predictors should be included in the model?

Generalized linear models are similar to linear, logistic, and Cox models in that much of the work in specifying and assessing the predictor side of the equation is the same no matter what distribution or link function is chosen. This can be especially helpful when analyzing a study with a variety of different outcomes, but similar questions as to what determines those outcomes. For example, in the depression example we might also be interested in cost, with a virtually identical model and set of predictors.

8.6 Further Notes and References

There are a number of book-length treatments of generalized linear models, including Dobson (2001) and McCullagh and Nelder (1989). In Chap. 7, we extended the logistic model to accommodate correlated data by the use of generalized estimating equations and by including random effects. The GLMs described in this chapter can similarly be extended and fit using the `xtgee` command in Stata and GENMOD procedure in SAS, which can be used with a variety of distributions. Random effects models can be estimated for a number distributions using the cross-sectional time-series commands in Stata (these commands are prefixed by `xt`) and with the NLMIXED procedure in SAS.

There are a number of approaches to modeling data with many zeros; Lachenbruch (2002) provides an accessible survey. He also considers the issue of power compared to simpler analyses. For example, in the simple two-group comparison of Sect. 8.3.1, we could use a nonparametric test like the Wilcoxon rank sum test. He shows that using two part or zero-inflated models, which explicitly model zeros, will often have higher power than simpler approaches that merely accommodate an outcome distribution with many zeros.

8.7 Problems

Problem 8.1. We made the point in Sect. 8.1.1 that a log transformation would not alleviate nonnormality. Yet we model the log of the mean response. Let us consider the differences.

(1) First consider the small data set consisting of 0, 1, 0, 3, 1. What is the mean? What is the log of the mean? What is the mean of the logs of each data point?
(2) Even if there are no zeros, these two operations are quite different. Consider the small data set consisting of 2, 3, 32, 7, 11. What is the log of the mean? What is the mean of the logs of the data? Why are they different?
(3) Repeat the above calculation, but using medians.

Problem 8.2. What would you need to add to model (8.5) to assess whether the effect of the treatment was different in whites as compared to non-whites?

Problem 8.3. Suppose the coefficient for $\hat{\beta}_2$ in (8.6) was -0.2. Provide an interpretation of the treatment effect.

Problem 8.4. For each of the following scenarios, describe the distribution of the outcome variable (Is it discrete or approximately continuous? Is it symmetric or skewed? Is it count data?) and which distribution(s) might be a logical choice for a GLM.

(1) A treatment program is tested for reducing drug use among the homeless. The outcome is injection drug use frequency in the past 90 days. The values range from 0 to 900 with an average of 120, a median of 90, and a standard deviation of 120. Predictors include treatment program, race (white/non-white), and sex.
(2) In a study of detection of abnormal heart sounds the values of brain natriuretic peptide (BNP) in the plasma are measured. The outcome, BNP, is sometimes used as a means of identifying patients who are likely to have signs and symptoms of heart failure. The BNP values ranged from 5 to 4,000 with an average of 450, a median of 150, and a standard deviation of 900. Predictors include whether an abnormal heart sound is heard, race (white/non-white), and sex.
(3) A clinical trial was conducted at four clinical centers to see if alendronate (a bone-strengthening medication) could prevent vertebral fractures in elderly women. The outcome is total number of vertebral fractures over the follow-up period (intended to be 5 years for each woman). Predictors include drug versus placebo, clinical center, and whether the woman had a previous fracture when enrolled in the study.

Problem 8.5. For each of the scenarios outlined in Problem 8.4, write down a preliminary model by specifying the assumed distribution, the link function, and how the predictors are assumed to be related to the mean.

8.8 Learning Objectives

(1) State the advantage of using a GLMs approach.
(2) Given an example, make reasonable choices for distributions, and link functions.
(3) Given output from a GLMs routine, state whether predictors are statistically significant and provide an interpretation of their estimated coefficients.