Journal of Educational and Behavioral Statistics
Summer 2002, Vol. 27, No. 2, pp. 147–161

Infinite Parameter Estimates in Logistic Regression: Opportunities, Not Problems

David Rindskopf
CUNY Graduate Center

Infinite parameter estimates in logistic regression are commonly thought of as a problem. This article shows that in principle an analyst should be happy to have an infinite slope in logistic regression, because it indicates that a predictor is perfect. Using simple approaches, hypothesis tests may be performed and confidence intervals calculated even when a slope is infinite. Some functions of parameters that are infinite are still well defined, and reasonable estimates of these quantities (in particular, LD50) may be obtained even when the maximum likelihood estimates do not, in a strict sense, exist. Surprisingly, these techniques can provide more reasonable and useful results than the most popular alternative method, exact logistic regression.

Keywords: *boundary estimates, exact statistics, inverted hypothesis test, logistic regression*

In certain cases, estimates of logistic regression coefficients can be infinite (either theoretically or practically, in the sense of being "very large"). One might think that this is a problem; indeed, some authors (e.g., Ryan, 1996) give examples and show that different standard computer programs can give strange (and very different) estimates of parameters for the same data set. Hosmer and Lemeshow (1989) also discuss this problem, and note that zero cells or complete separation will cause problems for the usual maximum likelihood (ML) estimation procedures. Mehta and Patel (1995) have presented examples using real data in which the traditional ML estimates do not exist, and provide an alternative approach using "exact" methods. An additional problem is that when the parameter estimates are large, the standard errors are also very large, and large sample theory does not appear to apply, leading to further complications in the interpretation of the results of the analysis.

In this article I show that the problems are, for the most part, illusory, and that differences among computer programs are not as large as might be believed. Further, I show that the problems that do arise can be dealt with in a straightforward way that can be implemented on some of the standard software packages now in use. Most of these techniques are "known," in the sense in that they are in the literature; in particular, the book *Statistical Modeling in GLIM* (Aitkin, Anderson, Francis, & Hinde, 1989) demonstrates a number of useful techniques. But the discussions cited here, in standard sources on logistic regression, about problems with infinite parameter estimates, lead me to believe that even many experts in the area are unaware of the possible solutions.

Binomial Example

I begin with a simple example of the problem and some possible solutions. Consider a binomial outcome; suppose one observes no failures in 10 trials. What inferences can be made? The usual large sample theory is not applicable here, but it would not be applicable with no (or all) failures even if there were 100 trials, or 1,000 trials. The estimate of the proportion is no problem; it is .00 (for the probability of failure) or 1.00 (for the probability of success). The problem is the standard error: Large-sample estimates would give 0 for the standard error, regardless of sample size. But intuitively it seems that an observed proportion of .00 is a more accurate estimate if the sample size is larger, and less accurate if the sample size is smaller.

Note that most small-sample methods are no better. For example, using a logit transformation, where $\text{logit}(p) = \ln[p/(1-p)]$, gives problems even with the estimate. An observed proportion of zero has a logit of minus infinity (–INF, as it will be denoted in the tables), and an observed proportion of one has a logit of plus infinity (+INF). Here is the first example of an infinite parameter estimate. The standard error is also +INF, so we not only have an apparently problematic estimate, but a useless standard error. The standard error is useless not only because it is infinite, and therefore can't be used in any standard formulas, but because we know that the implication of this is wrong: If a standard error indicates the precision with which a parameter is estimated, it can't be infinite if there is a reasonable amount of data. (Technically, in this case the problem is that the log-likelihood is not nearly quadratic, as it must be for the usual methods to apply; see Kalbfleisch, 1985).

One approach that has advantages in this case is Bayesian statistics. Nevertheless, I realize that most statisticians are not Bayesians, so I will concentrate on classical inferential procedures in this article. But one technique common in loglinear models, adding .5 (or some other small value) to cell counts when there are many zero counts in a table, is actually identical in effect to what a Bayesian would do if he or she had a slightly informative prior. For large, sparse, tables it is better not to add .5 to each cell; that corresponds to a stronger prior for such tables. Adding some small constant to each cell will make all parameter estimates finite, as well as standard errors, and will provide smoothed estimates of cell frequencies. On the other hand, it precludes 0 and 1 as plausible values for predicted proportions, and may smooth too much toward the equiprobability model. (For more refined methods of this sort, see Clogg, Rubin, Schenker, Schultz, & Weidman, 1991).

Confidence Intervals by Inverting a Hypothesis Test

If the usual large-sample and small-sample approaches are of no use in this situation, must we abandon classical statistical methods? My answer is no; we merely have to use some ideas that are seldom applied. The first is inverting a hypothesis test. In the case of $r = 0$ successes in $n = 10$ trials, for example, consider all the possible values of the population proportion π that would not be rejected by a hypothesis test at some level of significance α. These can all be considered plausible

values for the parameter. For zero failures in 10 trials, any value of π less than approximately .175 will give a likelihood ratio goodness-of-fit less than 3.84, so the 95% confidence interval for π is (.000, .175). See Aitkin, et. al., (1989, pp. 117–118) for a more detailed discussion of this example.

Figure 1 shows a plot of the likelihood-ratio goodness-of-fit statistic (GOF) for this example. The horizontal axis is the probability of success on a single trial; the vertical axis is –2 times the natural logarithm of the likelihood. At the value of the ML estimate ($\pi = 0$), the GOF is 0; at just under $\pi = .18$, the GOF is greater than 3.84, the critical value of the chi square distribution for a significance level of .05. Therefore, any value of π less than .18 is plausible, while values greater than .18 are implausible. (The more precise value $\pi = .175$ given in the previous paragraph can be obtained algebraically from the equation for the log-likelihood.)

Compare this interval with an alternative solution to this problem that involves the quasi-Bayesian idea mentioned above of adding .5 to successes and failures. This results in $r = .5$, $n = 11$. The estimate of the population proportion is now $p = .5/11 = 1/22 = 0.04545$. Using the logit transformation, finding a 95% confidence interval for the logit, and translating back into proportions, produces a confidence interval of (.006, .261) for the proportion (see Appendix for details). Adding .5 to r and 1 to n has changed the lower limit slightly (from 0 to .006), and changed the

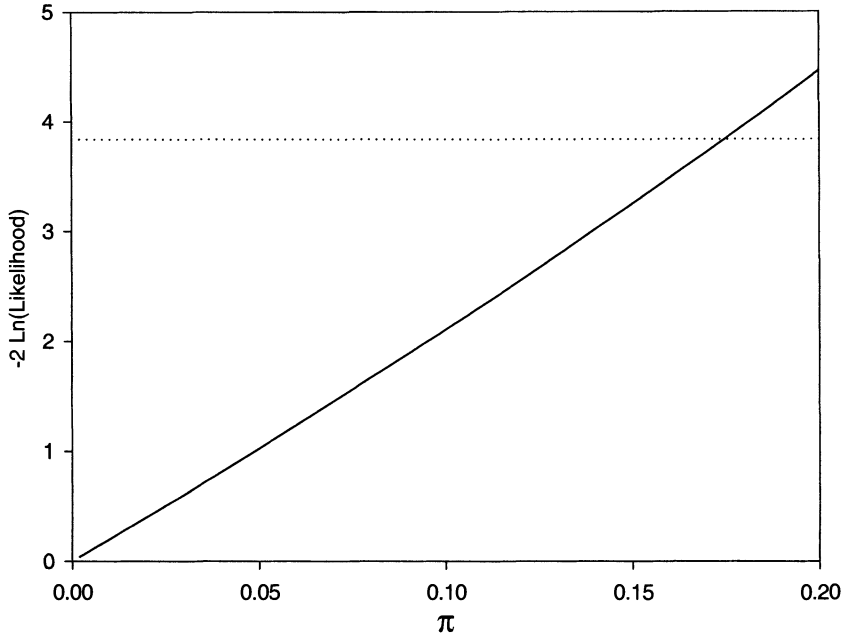

FIGURE 1. *Plot of –2 log likelihood for* r = 0, n = *10. Horizontal axis is probability, vertical axis is –2 times the natural logarithm of the likelihood.*

upper limit even more (from .175 to .261). Thus with fairly small n this procedure can have a larger effect than we might want.

Logistic Regression

One might wonder if the binomial example is related to the stated problem of infinite parameter estimates in logistic regression. In fact, the binomial model can be written as a special case of logistic regression with no predictor variables. The model is $\ln[\pi / (1 - \pi)] = \beta$. The confidence interval applied using the logit transformation in the previous section is equivalent to this formulation of logistic regression, and can be calculated using standard statistical packages (except for the final step). While most packages for logistic regression require a predictor variable, some can be tricked by creating a variable ONE = 1, and entering that as the only predictor. SPSS, for example, will detect that the variable ONE is collinear with the constant term; it will then omit the constant term and use ONE instead. If requested, SPSS will print exp(b), and 95% confidence intervals for exp(b). In this case, these are values for odds; to get proportions, divide the odds by 1 + odds.

Plotting Data for Logistic Regression.

We now proceed to examine a data set with a "real" independent variable. This data set is from Ryan (1996); see Table 1. While the relationship may not be obvious on inspection of the raw data, it becomes obvious when the data are plotted, as in Figure 2. (This data set is so simple that the plot shows what is happening. But in general, a useful plot for logistic regression with repeated X values, using SPSS, is to SORT by X; AGGREGATE, breaking on X; create an aggregated variable = MEAN(Y), where Y is the 0/1 outcome variable. For the aggregated data set, plot MEAN(Y) by X. If X has a large number of values, and few repeated X values, create a new X variable with 10 to 20 categories, depending on total sample size, and then use the same procedure.)

The plot here shows an obvious functional relationship, such that for any X less than or equal to 19, $Y = 0$, while for any X greater than or equal to 20, $Y = 1$. The

TABLE 1
Example Data Set from Ryan (1996, Table 9.1, p. 263), With Pairs of Values Representing X (Predictor) and Y (Outcome)

29, 1	25, 1	12, 0	24, 1	12, 0
26, 1	15, 0	15, 0	30, 1	20, 1
29, 1	16, 0	24, 1	25, 1	14, 0
26, 1	18, 0	24, 1	15, 0	27, 1
21, 1	15, 0	14, 0	29, 1	28, 1
24, 1	29, 1	26, 1	26, 1	23, 1
18, 0	22, 1	27, 1	19, 0	16, 0
17, 0	22, 1	16, 0	17, 0	24, 1
26, 1	10, 0	19, 0	11, 0	19, 0
22, 1	10, 0	18, 0	13, 0	21, 1

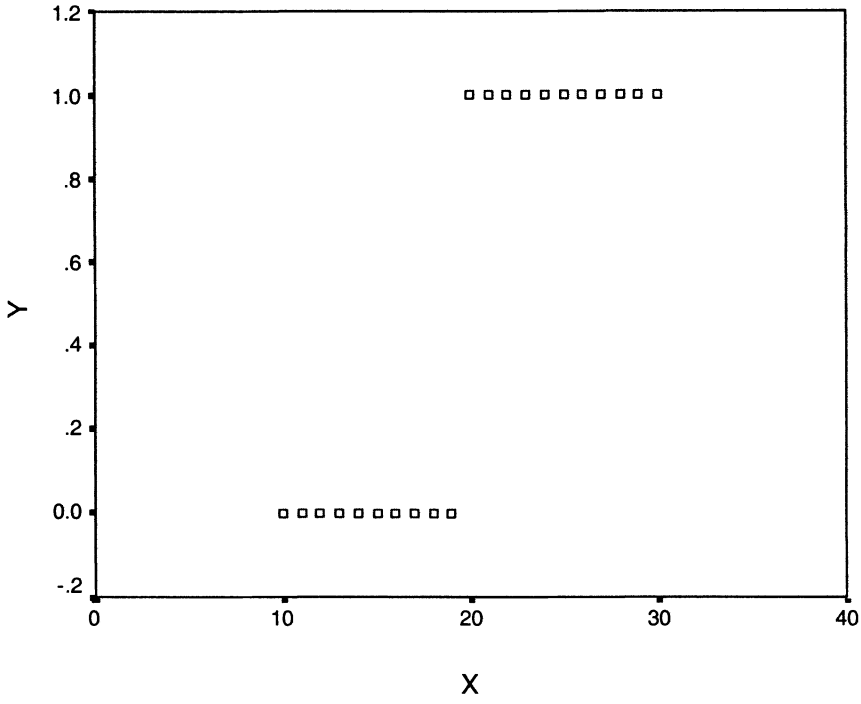

FIGURE 2. *Plot of artificial data from Ryan (1996).*

rule is simple to state verbally, but the perfect relationship presents an obvious problem for logistic regression, which was not meant to represent perfect relationships. This seems very strange: We want to predict well, but when we predict too well, our tools are inadequate. Here we can see two problems for logistic regression: (a) the slope is infinite, and (b) the intercept is indeterminate.

Parameter Estimates, Interpretation, and Use

Table 2 shows the population values from which the data were generated, and some solutions when logistic regression is applied to these data. Some solutions are from Ryan (1996), and others are from ML programs. The second S-Plus results are from a recoding in which 20 was subtracted from each X value, so that the new X variable was approximately centered. The estimates of the intercept and slope shown in Table 2 vary widely, leading to a hypothesis that nothing useful can come from these analyses. (Of course, due to the transformation of X in the last method, the intercept will be different.) Since we already pointed out that the slope is actually infinite, this would be a reasonable conjecture. But in fact several useful results can be deduced, and these can lead us to conclude that some of the solutions in the table are reasonable, while others are unreasonable (yes, there are reasonable and unreasonable estimates of parameters that are infinite!).

TABLE 2
Logistic Regression Estimates for Ryan Data Set, Using Different Methods of Estimation and Computer Programs

Program	b_0	b_1	$LD50 = -b_0/b_1$
Ryan: "Population"	−78.8	4.0	19.70
Ryan: Initial	−25.3270	1.2642	20.03
Ryan: Exact	−12.4801	1.7414	7.17
S-Plus: GLM	−243.57	12.46	19.55
S-Plus: X-20	5.638	17.4605	19.68 *

Note. *Adjusted back to original scale of *X*.

First, consider that the usual intercept and slope form for the equation is not the only way to specify the model; sometimes other ways that use different parameters are of more interest. For example, in many circumstances in biology and medicine, the slope and the so-called LD50 are the important parameters. LD50 is the value of the predictor, *X*, for which the response rate is 50%. The same construct is used in item response models in psychometrics: The item difficulty is usually specified as the value of theta, the underlying trait, for which the probability of answering an item correctly is .50. Thus, most IRT models write the logistic (or probit) equation not in terms of an intercept and slope, but in terms of the LD50 (for Rasch models), or the LD50 and the slope (for two-parameter models).

In terms of the usual logistic regression parameters, $LD50 = -b_0/b_1$; that is, the negative of the intercept divided by the slope. One might think that if the slope is going towards infinity, and the intercept seems fairly indeterminate as well, then what hope is there for the quotient to be of any use? However, this intuition turns out not to be accurate. A better intuition comes from viewing the plot of such a data set. A logistic curve must rise from a floor at or near zero for *X* values less than 19, up to a ceiling at or near 1 for *X* values greater than or equal to 20. The point at which the curve crosses .5 on the vertical axis (LD50) clearly should be about halfway between *X* values of 19 and 20. Figure 3 illustrates these facts using a logistic curve with a slope of 10; while quite far from being infinite, that value for the slope produces quite a steep curve over the range of *X* values of interest.

It is not surprising, therefore, that Table 2 shows, for most of the logistic regression results, that the LD50 is about half way between 19 and 20, as it should be for these data. (The LD50 value for the last method was translated back to the original scale for comparability with the other results.) Ryan's initial estimates give an LD50 slightly higher than 20, but this is reasonably close for an initial estimate of parameters. Even though one parameter is going toward +INF, and another toward −INF, their ratio is going toward a finite (and useful) limit. Only the exact method gives a result that is clearly wrong. We conclude that if *X* is continuous, the usual logistic regression can provide a reasonable estimate of LD50 even in the extreme case of an infinite slope. Further, the "exact" method, which because of its name

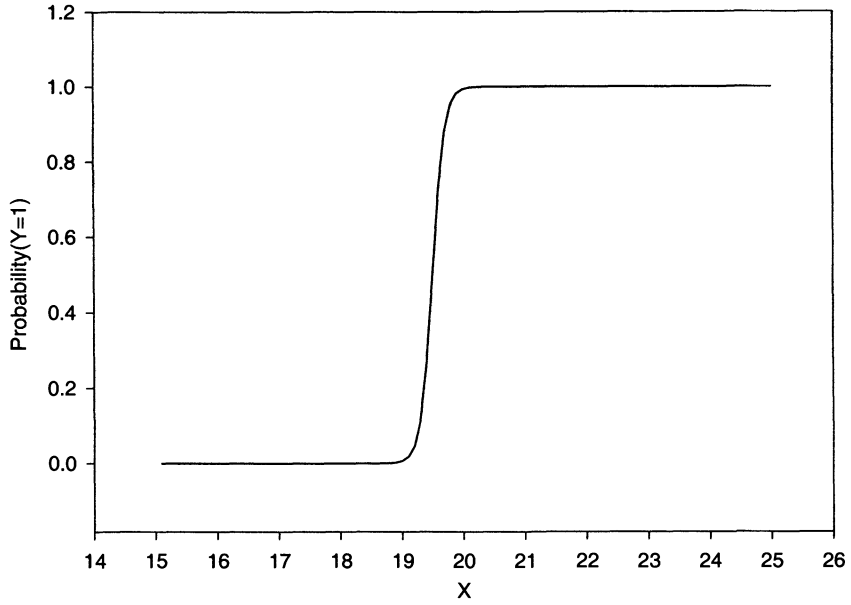

FIGURE 3. *Plot of logistic curve with slope equal to 10.*
Note: This curve provides a nearly perfect fit to the Ryan (1996) data.

seems to have magical appeal, can give results that are problematic. We give further examples of these problems later.

Note the results (except for the "exact" analysis) do not disagree as much as implied by Ryan. The LD50 seems right for all but the "exact" method. Further, if one calculates predicted proportions for all X values in the observed range, all except the "exact" method give reasonable predictions. That is, for X values less than or equal to 19, most of these equations predict nearly .00, and for X values greater than or equal to 20, they predict nearly 1.00, as they should. Therefore, even though the slope and intercept are headed towards infinity in opposite directions, the linear combinations produce predicted logits (and ultimately predicted proportions) that seem correct for each of the ML estimation procedures. The "exact" method fails here, giving predictions that are far from the observed data; the equation gives predicted probabilities near one for most of the range of observed X values, not just those over 20.

With more than one predictor, there is no standard method of calculating LD50. Some care is needed, because dividing the intercept by the slope for any predictor will give the LD50 for that variable when all other variables have the value zero. In most cases, this will not be what is desired. One solution is to center all predictors before running the logistic regression. The LD50 for each predictor will then be meaningful in the context of a zero value for all other predictors. For categorical

predictors, coding is still important, but (unlike the case with continuous predictors) most standard methods of coding will produce reasonable results.

In this data set the slope is obviously infinite; the standard errors (not reported here) will also tend toward infinity. Is there a reasonable way to test whether the slope is significantly different from zero? Yes; merely fit the data with only a constant term in it, using the method described above. If the value of –2 log-likelihood changes by more than 3.84, the omitted variable was a useful predictor; otherwise, it is nonsignificant. While this method can be used with nearly any logistic regression program, to get a confidence interval for the slope requires tools not always found in standard programs. We will illustrate these tools in the discussion of the next example.

Example from Mehta and Patel (1995): CD4, CD8, and HIV

Mehta and Patel (1995) analyzed the following data set in their groundbreaking article on exact logistic regression. The data in Table 3 shows the relationship between two biochemical measures, CD4 and CD8, and the development of HIV. The data set is small, and ML methods might be suspected to have problems. In fact, Mehta and Patel say that ". . . we cannot estimate the [logistic] parameters . . . by the maximum likelihood method because the observed data fall on the boundary of the parameter space . . ." The LogXact demonstration program (Cytel Software Corporation, 1992) states that ". . . the maximum likelihood estimates do not exist and no convergence is possible for this small data set." However, as we will see, these problems can generally be resolved satisfactorily, while the "exact" method may give some misleading results.

Descriptive Statistics for HIV Data

In this example, we are trying to predict HIV status in children from CD4 and CD8 cell counts, which have been categorized on a three-point scale (i.e., into three categories). This data set is discussed in Mehta and Patel (1995), and it is used in the

TABLE 3
Relationship of CD4 and CD8 to Development of HIV in Children

		Develop HIV?			
CD4	CD8	Yes	No	Total	P(yes)
0	2	1	0	1	1.0000
1	2	2	0	2	1.0000
0	0	4	3	7	.5714
1	1	4	8	12	.3333
2	2	1	2	3	.3333
1	0	2	5	7	.2857
2	0	0	2	2	.0000
2	1	0	13	13	.0000

demonstration of LogXact. In that demonstration, the two predictors are coded using two dummy variables for each, with the last level (level 2) of each used as the base category. This is the worst possible choice for producing interpretable parameters, as can be seen from the results in Table 4, which shows the data arranged in a two-way table. On the other hand, it is the "best" choice for producing problematic parameter estimates, which was apparently the goal of the demonstration. The lower right-hand corner of each section of Table 4 shows the value that will determine the intercept (i.e., ln(1/2) = –.69). The other cells in the bottom row and right-hand column are all proportions of either zero or one, so the estimates of the parameters corresponding to the dummy variables will all be either minus or plus infinity. For the remaining cells, it is the sums of various parameters that matter; these can and should converge to reasonable values even though the parameters contributing to these sums go towards infinity (in opposite directions).

Parameter Estimates, Interpretation, and Use: HIV Data

Table 5 shows the parameter estimates from two computer programs. One is a program that produces "exact" analyses; the second is an ordinary logit model analysis program that uses maximum likelihood estimation; the third is ML using S-Plus **glm.** The ML parameter estimates look odd, but they are approaching the right answer more closely than the "exact" parameter estimates: The CD4 parameters are heading towards infinity, and the CD8 parameters towards negative infinity. For a logit model, parameters with an absolute value of about 20 for a

TABLE 4
HIV Data Arranged as a 3 × 3 Table

	CD8		
CD4	0	1	2
Proportions with HIV			
0	.57	*	1.00
1	.29	.33	1.00
2	.00	.00	.33
Odds of HIV			
0	1.333	*	+INF
1	.400	.500	+INF
2	–INF	–INF	.500
Logit [= ln(odds)] of HIV			
0	.288	*	+INF
1	–.916	–.693	+INF
2	–INF	–INF	–.693

Note. *No observations in this cell.
+INF = plus infinity, –INF = minus infinity.

TABLE 5
Parameter Estimates for Original Coding of HIV Data

Parameter	LogXact	ML*	ML: S-Plus
Intercept	−.6931	−.6931	−.6931
CD4_0	2.9353	20.958	8.958
CD4_1	2.4456	19.754	7.754
CD8_0	−2.2471	−19.977	−7.977
CD8_1	−2.3190	−19.754	−7.754

Note. *ML estimates produced using MQUAL, written by D. Rindskopf in APL.

dummy variable are close enough to infinity for practical purposes (e.g., a logit of 20, represents odds of nearly 500,000,000, and a proportion of 0.9999999979388; a slope of 20 changes the odds by a factor of 500,000,000). A slope parameter of eight changes the odds by a factor of 2,980, and a logit of eight corresponds to a proportion of .9997.

We also want to examine the expected proportions produced by these solutions. Of course the expected logits (which are translated into expected odds, and then into expected proportions) are functions of the parameter estimates. Because of this, intuition would lead us to suspect that we cannot get useful estimates, but again intuition is wrong. Table 6 contains the results for ML estimation; Table 6 also contains the results for the "exact" method. The ML estimates of cell proportions accurately reproduce the observed values, both for the cells that have "extreme" observed proportions of zero and one, as well as the other cells of the table. Therefore, the main-effects-only model fits the data well. Furthermore, the ML procedure produces an estimate for the cell with no observations. The "exact" parameters, on the other hand, come nowhere near reproducing the observed proportions of

TABLE 6
Expected Proportions for HIV Data, Original Coding

	CD8		
CD4	0	1	2
Expected proportions from ML Estimates			
0	.571	.625	1.00
1	.286	.333	1.00
2	.000	.000	.333
Expected proportions from the Exact Estimates			
0	.499	.481	.904
1	.379	.362	.852
2	.050	.047	.333

zero and one in the data, nor do they accurately reproduce the three nonzero cells not in the last row or column.

Now we provide another illustration that functions of parameters may be useful even when the parameter estimates are heading towards infinity. Though the coefficients for both CD4_0 and CD4_1 are very large positive numbers, they are very close to each other; similarly, the coefficients for CD8_0 and CD8_1, though large and negative, are close to each other. We might therefore hypothesize that each pair of coefficients is equal, and rewrite the model with two predictors instead of four. We define a dummy variable *CD4A* to equal 1 if CD4 is either 0 or 1, and to equal 0 if CD4 is 2; similarly, we define a dummy variable *CD8A* to equal 1 if CD8 is either 0 or 1, and to equal 0 if CD8 is 2. (This is the same as adding together the corresponding original dummy variables to impose a restriction, as discussed in Rindskopf, 1984.) The resulting model has a likelihood-ratio chi-square of 1.434 ($df = 5$, $p = .921$), and therefore fits the data quite well. We can expect that the resulting estimated probabilities are more stable than those from the previous model. As might be expected, the parameter estimates are still approaching infinity, as are the standard errors, but this does not affect the model fit, the interpretation, or the predicted proportions.

Significance Testing Using Differences in Model Fit

As might be expected, not only are the ML parameter estimates extremely large, but so are their standard errors. For the HIV data set, the standard errors were 3,425 for each parameter except the intercept, the point at which the procedure terminated. The usual methods for testing the significance of parameters, or for finding confidence intervals for the parameters, are therefore of no use. However, we may easily test for significance of parameters and sets of parameters using the difference in model fit statistics, much as in the usual procedure for loglinear and logistic regression models (and analogous to tests of R-square change in regression). Finding confidence intervals is more involved, and cannot be done using some software (e.g., SPSS logistic regression), but can be done using any software that allows *offsets* to fix (logistic) regression coefficients (e.g., SAS GENMOD, SPSS loglinear, GLIM, and S-Plus).

First we test whether each parameter is significantly different from zero. To do this, we compare the fit of the full model (here the main effects model) to models with each of the parameters omitted in turn. The main effects model fits perfectly here; the likelihood ratio (LR) chi-square is nearly zero. Table 7 shows the LR for models with each parameter omitted; ordinarily, we would subtract the LR for the full model from these values, but as it is zero, the differences are the same as the fit statistics. Clearly, each of the parameters differs significantly from zero. As will be seen in the examination of confidence intervals, the results for the "exact" method are somewhat different: CD8_1 is barely significant at the .05 level, and CD8_0 is not.

Next we do a test that is not really needed here, given the previous results, but which we include as an illustration. To test whether CD4 status is an important pre-

TABLE 7
HIV Data: Tests That Parameters Equal Zero

Effect Omitted	LR difference	p*
CD4_0	11.536	.0007
CD4_1	11.018	.0009
CD8_0	8.010	.005
CD8_1	8.188	.004

Note. *Each test has one *df.*

dictor, we fit a model with both CD4 dummy variables omitted and compare the fit to that of the full model; the importance of CD8 is tested in the same way. Because two parameters are removed for each test, the LR difference is referred to as a chi-square distribution with two degrees of freedom. Table 8 contains the results; as would be expected, each variable is important. Comparable tests using the "exact" procedure produce similar results here: Both sets of parameters are highly significant. This is somewhat surprising given the borderline significance of the individual CD8 parameters using the "exact" method.

TABLE 8
HIV Data: Tests That Sets of Parameters Equal Zero

Effects Omitted	LR difference	p*
CD4_0 and CD4_1	13.760	.001
CD8_0 and CD8_1	9.187	.010

Note. *Each test has two *df.*

Confidence Intervals Using Hypothesis Test Inversion

Finding confidence intervals is similar to testing hypotheses that parameters equal zero, in that the procedure involves a comparison of models. For confidence intervals, we must find values of a parameter that change the model fit by 3.84 (the critical value of chi-square with one degree of freedom). Computer programs such as SAS GENMOD, SPSS loglinear, GLIM, and S-Plus allow the specification of an "offset", which can be used to accomplish this task. Here I will demonstrate the use of the S-Plus **glm** procedure. The usual model is specified using the command

```
glm(hiv ~ cd4_0 + cd4_1 + cd8_0 + cd8_1, family = binomial(link = logit))
```

To restrict the coefficient for CD4_0, for example, to equal 2, we would change the model specification to:

```
hiv ~ cd4_1 + cd8_0 + cd8_1 + offset(2* cd4_0)
```

By trying different constants, one can rather quickly find the value that changes the residual deviance (as it is called in the output) by 3.84. The similarity of this method to the hypothesis test is that, for hypothesis testing, one can fix the parameter at zero by removing the predictor. That method, of course, can be used with any standard logistic regression package, while offsets are available only in some packages.

Table 9 shows the 95% confidence intervals produced by the "exact" method, and by ML estimation using the method of inversion of the hypothesis test. Notice that the "exact" values are much more conservative than the ML estimates for the confidence intervals. For the CD8 dummy variables, one "exact" confidence interval does not cover zero, and the other barely does, while the ML estimates clearly exclude zero as plausible values.

One may also be interested in confidence intervals around the predicted proportions for some (or all) cells in the table, or more generally for some particular value X_i of a set of predictor variables X in a logistic regression. Generally one could get these by using the variances and covariances of the parameter estimates, but because these are problematic in this situation, a different approach is required. Using the method of inverting the hypothesis test, we obtained confidence intervals for all model parameters, including the intercept. As pointed out by Mehta and Patel (1995), one can make the intercept represent the predicted logit for any particular value X_i of the predictors by transforming the predictors using $X^* = X - X_i$. Using X^* in the logistic regression, the intercept will now represent the logit for $X = X_i$, and the confidence interval for the logit can be transformed to a confidence interval for the odds or proportion expected to respond at that value of X.

This example demonstrates that although ML estimates may technically be undefined in some cases, we can still follow sensible procedures to get parameter estimates, useful functions of parameter estimates, and confidence intervals. At the same time, "exact" methods may produce parameter estimates that are clearly deficient, functions of parameter estimates that are also deficient, and confidence intervals that are too conservative.

TABLE 9
"Exact" and ML Confidence Intervals for HIV Data

	LogXact		ML*	
Parameter	Lower	Upper	Lower	Upper
CD4_0	.54	+INF	2.12	+INF
CD4_1	.45	+INF	1.50	+INF
CD8_0	–INF	.07	–INF	–1.28
CD8_1	–INF	–.02	–INF	–1.28

Note. *Produced using S-Plus 4.5, using offset () in equation.

Discussion and Conclusion

While SPSS logistic regression and S-Plus **glm** will produce parameter estimates even when one or more parameters is going to infinity, some programs may not do so. Whether a program believes it has "converged" may be monitored in more than one way. If a program is monitoring changes in the parameter values, then it will never believe the procedure has converged, because they will keep getting bigger (towards +INF) or smaller (towards –INF). On the other hand, if the program is monitoring changes in the log likelihood, then it will converge at some point, as the log likelihood is bounded. In either case, the decision about whether the results will be printed lies with the programmer, and practices may differ from one package to another.

Note that it is not useful to extend the number system to define a parameter as being +INF or –INF. If this is done, then linear combinations are not defined, so all of the derived quantities of interest that were demonstrated here could not be computed. While it is true that these would be the most accurate estimate of the parameters (they are, in fact, infinite), it makes them worthless. Therefore, in these cases, the parameter estimates we use cannot be unique; they depend on the stopping criterion. Two computer programs (or one program with the criterion reset) will give different (finite) estimates of the parameters that are infinite. However, they should give (nearly) identical estimates of all important functions of the parameter estimates, as demonstrated above. It may seem strange to say that a parameter estimate is useful only if it is not unique; this contradicts the typical desiderata for estimation. However, by giving up uniqueness in estimating the model parameters, we can obtain quite useful results, and give up only a small amount in that we can estimate important quantities almost, but not quite, uniquely.

Why does the "exact" method fail to give reasonable results? I can only conjecture about this. When a parameter is infinite, then the exact method should give an infinite estimate also. The LogXact program uses what Mehta and Patel (1995) call a "median unbiased estimate." While such an estimate may have some desirable properties, it apparently has other properties that are problematic, leading to inaccurate estimates of many important functions of the parameters.

In conclusion, I hope that this article changes the common view of infinite parameter estimates in logistic regression. Infinite parameter estimates are desirable in logistic regression, because it means that prediction is perfect (the presumed goal). Unfortunately, logistic regression is not the perfect mathematical model in this situation; it is attempting to approximate a step function, the actual model for perfect prediction. In this case, as well as near-perfect prediction, a reasonable alternative to the usual large-sample standard errors, confidence intervals, and hypothesis tests is to establish confidence intervals by inverting the usual chi-square tests, or to test hypotheses about coefficients or sets of coefficients using the usual technique of comparing model fits. Furthermore, many useful functions of the parameters are accurately estimated even when one or more parameters is approaching infinity. While "exact" methods may have advantages in some cases, particularly when

sample sizes are small, they are not a panacea, and can be problematic in cases where ML estimates are not.

Appendix

Calculations of 95% confidence interval for a proportion, using logit transform; $r = 0$, $n = 10$, with .5 added to r and $n–r$.

```
p = (r+.5)/(n+1)
Odds = p/(1-p) = 0.047619.
Logit = ln(Odds) = -3.0445
SE = sqrt(1/[n*p*(1-p)]) = 1.023532
W = 1.96*SE = 2.006
LowerLogit = Logit - W = -5.0505
LowerOdds = exp(LowerLogit) = 0.00640
LowerProp = LowerOdds /(1 + LowerOdds) = 0.00636
```

Similar calculations produce the following results for the upper limit:

```
UpperLogit = -1.038522
UpperOdds = 0.353977
UpperProp = 0.2614
```

References

Aitkin, M., Anderson, D., Francis, B., & Hinde, J. (1989). *Statistical modeling in GLIM.* Oxford, UK: Oxford University Press.

Clogg, C. C., Rubin, D. B., Schenker, N., Schultz, B., & Weidman, L. (1991). Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression. *Journal of the American Statistical Association, 86,* 68–78.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Kalbfleisch, J. G. (1985). (2nd Ed.). *Probability and statistical inference, Volume 2: Statistical inference.* New York: Springer-Verlag.

LogXact. (1992). *Software for exact logistic regression.* [Computer program] Cambridge, MA: Cytel Software Corporation.

Mehta, C. R., & Patel, N. R. (1995). Exact logistic regression: Theory and examples. *Statistics in Medicine, 14,* 2143–2160.

Rindskopf, D. (1984). Linear equality restrictions in regression and loglinear models. *Psychological Bulletin, 96,* 597–603.

Ryan, T. P. (1996). *Modern regression methods.* New York: Wiley.

Author

DAVID RINDSKOPF is Professor, Educational Psychology and Psychiatry, CUNY Graduate Center, 365 Fifth Ave, New York, NY 10016; drindskopf@gc.cuny.edu. He specializes in categorical data, missing data, multilevel models, and structural equation models.

Manuscript Received March 2000
Revision Received January 2002
Accepted February 2002