Interpreting the coefficients of loglinear models using relative risk or incidence rate ratio.

1) Starting point.

Let's say we have a simple model,

1a) $Log(U) = \text{Const} + B_1X_1 + B_2X_2 + ...$

Where the B's are model coefficients, and the X's are the variables (usually dummy variables) and the U are predicted counts.

When X_1=0, we have:

1b) $Log(U) = \text{Const} + 0 + B_2X_2 + ...$

and when X_1=1

We have

1c) $Log(U) = \text{Const} + B_1 + B_2X_2 + ...$

So we can always say, as a simple function, that the coefficient B_1 represents an increase in the log of predicted counts. If B_1=2, for instance, we could say that 'this model shows that factor X_1 increases the predicted log count by 2 (all other factors held constant)' because equation (1c)- equation (1b)= B_1. This is true but not the most illuminating thing to say.

Remembering that e^0=1, we can exponentiate equation (1b) to get

1d) $U = e^{Const}e^{B_2X_2}$

and when X_1=1, we can exponentiate equation (1c) to get

1e) $U = e^{Const}e^{B_1}e^{B_2X_2}$

If we take the ratio of (1e)/(1d), we get e^{B1}. If for the sake of discussion we give B_1 the arbitrary value of 2, e^2=7.4, we could say that 'variable X_1 increases the predicted counts by a factor of 7.4' (all other factors being held constant). The cells that have X_1=1 have predicted counts, or an incidence rate ratio (using Stata's terminology) or a relative risk (using Agresti's terminology) of 7.4 compared to the cells that have X_1=0. Think of equation (1d) as the predicted value of events or counts for one cell (where X_1=0) in a big table, and equation (1e) as the predicted valued of events or counts for another cell where dummy variable X_1 takes on a value of 1.

----

2) See also my "Notes on different families of distributions" document for a section on why loglinear model coefficients are also (in some cases) log odds ratios.