**Appendix: Statistical Methods for Estimating the Probability of Causation**

We assume a simple model where the time to an episode of pharyngitis while positive for Group A strep (negative for Group A strep) is given by an exponential distribution with parameter  (). We assume this model applies within strata defined by the study year (March 2002 to March 2003 versus April 2003 to March 2004) and High or Low colonization rates (students whose mean colonization rate exceeds the overall median colonization rate versus students whose mean colonization rate is less than the overall median colonization rate). We further assume that the proportion of follow-up time that the students within a stratum are positive for Group A strep is given by p. The data within a stratum can be classified in the following table where the N carriage episodes are categorized as being positive or negative for GAS, as are the Y episodes of pharyngitis.

**Table A1: Results of GAS testing from the monthly carriage data and all reported pharyngitis cases for a generic stratum defined by year of follow-up and overall colonization rate.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Group A StrepNegative  | Group A StrepPositive  | Total |
| Carriage Visits | N- | N+  | N |
| Pharyngitis Visits  | Y- | Y+ | Y |

Using properties of the exponential distribution, one can show that Y-  follows a Poisson distribution with parameter T (1- p). Similarly, Y+ follows a Poisson distribution with parameter  T p. Finally, N-  follows a binomial distribution with parameters N and p. It follows that the log-likelihood (up to a constant) is given by

 Y+ log T p) - T p + Y- log{ T (1- p)} -  T(1-p) + N+ log(p) + N- log(1-p),

where T is the total years of follow-up for everyone. Maximization of this function produces the maximum likelihood estimates:

  = Y+ /(T p)

 = Y- /{T (1-p)}

 p = N+ /N,

where with a slight abuse of notation, we make no distinction between a parameter and its estimate.

A quantity of interest is the probability of causation or PC see [1,2]. For a patient presenting with GAS and pharyngitis, PC can be interpreted as the probability that GAS caused pharyngitis. Equivalently, one can view PC as the probability that GAS sensitive antibiotics would cure the pharyngitis in such a patient. One can show that under the exponential assumption and under the assumption that GAS cannot *prevent* pharyngitis in a patient who would otherwise have pharyngitis, and that (within a stratum) GAS colonization is effectively assigned at random one obtains

 PC = 1 - /- )/

Using the notation of Table 1, we can express our estimate of PC as

 PC = 1 - Y- N+ /Y+ N-

We form PCs for each stratum, say PC(1,H), PC(2,H), PC(1,L), and PC(2,L), where, for example, PC(2,L) is the probability of causation for study year 2 for students with low colonization rates. We obtain an overall PC for students with low colonization rates by forming

PC(L) = { PC(1,L) w(1,L)+PC(2,L)w(2,L)}/{w(1,L)+w(2,L)},

where w(j,L) is the number of cases of pharyngitis in study year j=1,2 for students with low colonization rates. The construction of PC(H) is analogous. To form an overall probability of causation we form

PC = { PC(1,L) w(1,L)+PC(2,L)w(2,L) + PC(1,H) w(1,H)+PC(2,H)w(2,H) }/{ w(1,L)+w(2,L) w(1,H)+w(2,H) }

where w(j,H) is the number of cases of pharyngitis in study year j=1,2 for students with high colonization rates. Under our assumptions, both the stratum specific and weighted PCs can be interpreted as the probability that effective antibiotics will cure the infection for a student presenting with group A strep and pharyngitis. We obtain estimates of PC(L), PC(H),and PC, by replacing PC(j,k) with its estimated value, where j=1,2 and k=L,H. For cells with counts of zero, we substitute the value .5. Estimates of PCs less than 0 were replaced by zero.

The methods described above can equally be applied to other forms of strep such as Group C strep, Group G strep, or any strep. The nonparametric bootstrap, where we randomly sample students with replacement, can be used to obtain confidence intervals for all parameter estimates.

**REFERENCES**

1. Levin ML (1953). The occurrence of lung cancer in man. *Acta Unio Internationalis contra Cancrum* 9 531-541

2. Greenland S (1999). Relation of probability of causation to relative risk and doubling dose: A methodologic error that has become a social problem. *American Journal of Public Health* 89 (8) 1166-1169.

3. Efron B, Tibshirani RJ (1998). An Introduction to the Bootstrap. New York, Chapman & Hall.