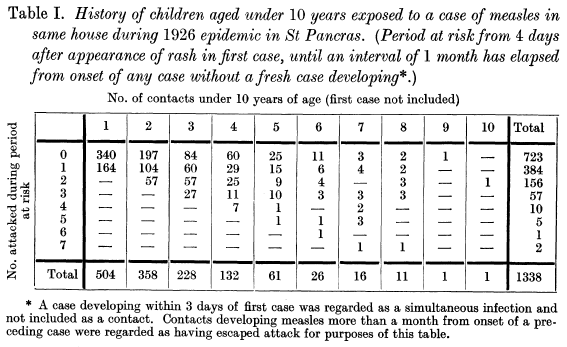
**Names:**

**EMD 538 Problem Set #1 – Chain Binomial Models**

Consider the data presented by Greenwood (1931) on the distribution of secondary measles cases among households with 1 index case and *m* = 4 contacts.



1. What is the secondary attack rate (SAR) for the households with *m* = 4 contacts, assuming all infected contacts had symptom onset within the maximum serial interval from the primary case?
2. What is the expected distribution of secondary infections within households of size *m* = 4 if there is NO ongoing transmission and *p* = 0.3?
3. What are the possible chains that could result in all 4 contacts being infected? What is the probability of each chain (in terms of *p* and *q*) under the Greenwood assumption?
4. Greenwood presents his expected distribution of secondary measles cases from the chain binomial model with *m* = 4 in Table III. What values of *p* and *q* did he use to arrive at this expected distribution? Show how he derived these estimates.
5. Why might the values of *p* and *q* for households with *m* = 4 be expected to differ from those for households with *m* = 3 in this study? **Give at least two reasons.**

(HINT: I don’t want to know why the estimates might differ, but rather why the “true” values might differ. Think about what assumptions you are making and why they might be violated.)

1. What would be the expected distribution of secondary cases for *m* = 4 under the Reed-Frost assumption given the values of *p* and *q* you calculated in question 4?
2. **EXTRA CREDIT:** Which assumption (Greenwood vs Reed-Frost) provides a better fit to the observed distribution of secondary cases when *m* = 4? Show the statistical criteria you used to determine this.