Models In Epidemiology And Biostatistics
Gordon Hilton Fick

Sample Size Determination
Based on Confidence Intervals
for the Odds Ratio

A Single Two By Two Table
The Confidence Interval for the Odds Ratio

We will describe the study as cohort [prospective] so that disease status is the outcome and exposure is explanatory.
We will assume that exposure increases the probability of disease.
Then the OR is presumed greater than 1.

The confidence interval for OR is not symmetrical about the estimated OR.
Then the lower limit of the confidence interval is of interest.
The distance between the lower limit and the estimated OR is the lower width.

The sample size formula requires the lower width relative to the OR; called the relative lower width (R)

For a case-control study, one interchanges disease and exposure in the expressions.

The formula

$$n_0 > \frac{z^2_{\alpha/2}}{\log^2(1-R)}\left(\frac{1}{p_0(1-p_0)} + \frac{1}{kp_1(1-p_1)}\right)$$

This formula can be derived using the delta method. A derivation is included in a separate file.

Five Pieces of Information

R : the relative lower width: OR – LL = R*OR

k : proportion exposed to unexposed $\frac{n_1}{n_0}$

$1-\alpha$: usually 95% so that $z_{\alpha/2}=1.96$

p_0 : the probability of disease among the unexposed

OR : odds ratio [used to compute p_1 in the formula]

How to select R ?

- Perhaps one would wish to have LL greater than 1. This means that one would select R to be less than : (OR - 1) / OR

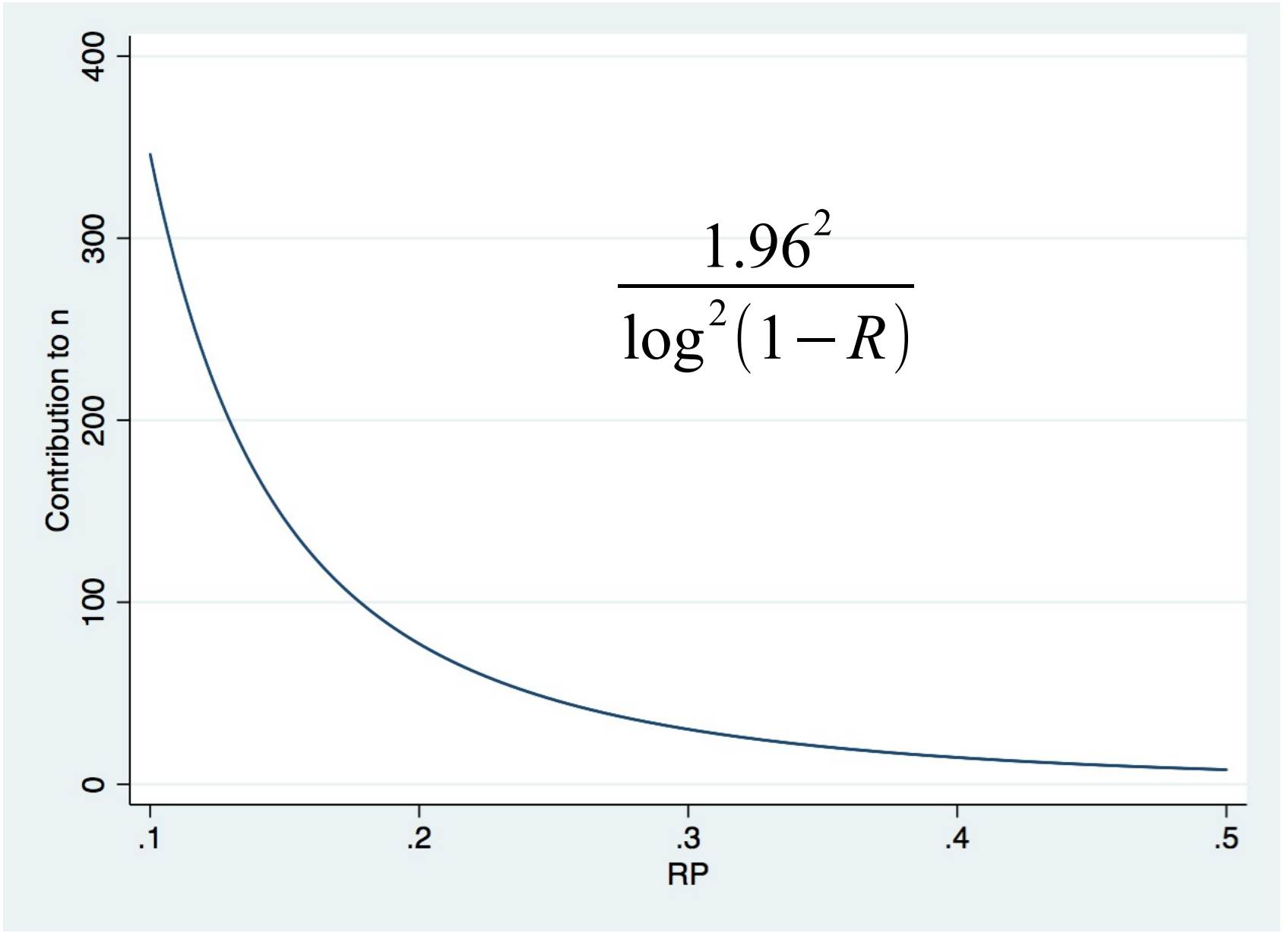
The formula can be seen as a product of two terms

The first term is : $\frac{1.96^2}{\log^2(1-R)}$

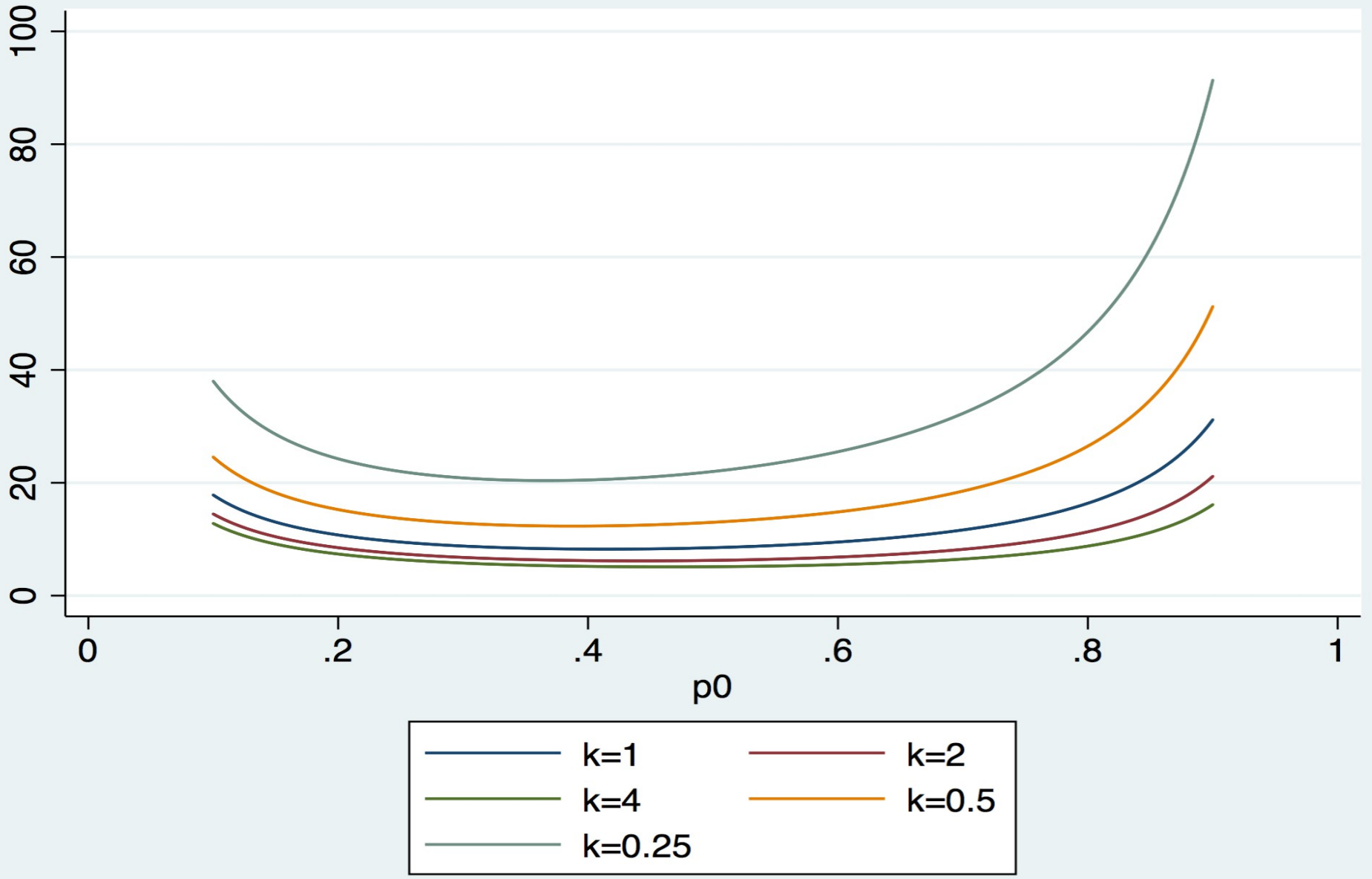
The second term is:

$$\left(\frac{1}{p_0(1-p_0)}+\frac{1}{kp_1(1-p_1)}\right)$$

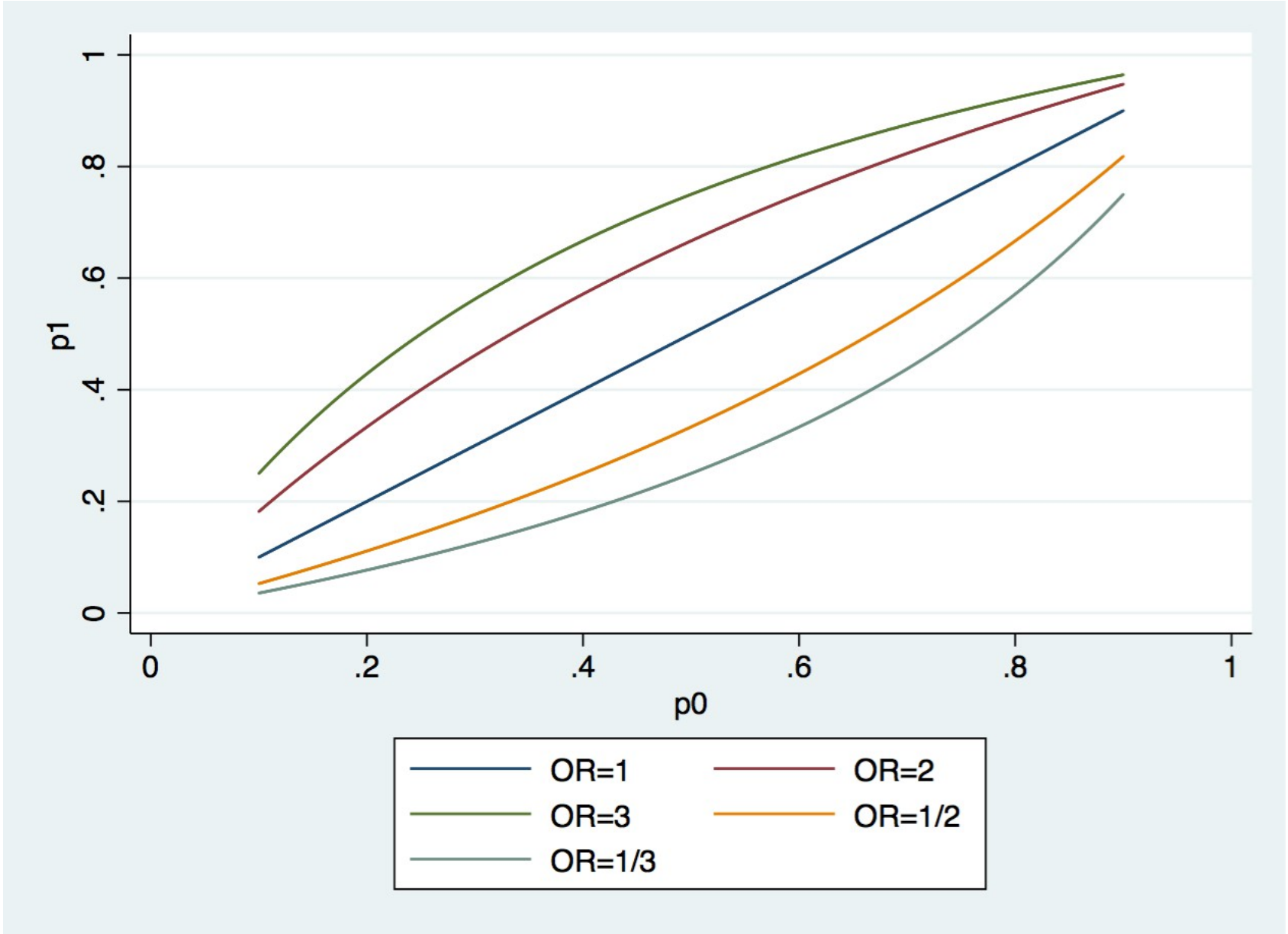
The first term versus R



The second term with OR=2



Determining p_1 from OR and p_0



A Single Potential Dichotomous Modifier/Confounder

Two 2 by 2 tables

Sample Size Determination 'needed' for each table

If the crude table is used, overall sample size will be adequate [refer to single table]
Similarly if one is able to use an 'adjusted' OR estimate.

Two Potential Dichotomous Modifiers/Confounders

Four 2 by 2 tables

Sample Size Determination 'needed' for all four of these tables

Then, One-At-A-Time tables will have adequate sample sizes as will the Crude table

Now, suppose there are k potential dichotomous modifiers/confounders

Now, we have 2^k two by two tables and, in principle, each table 'needs' a sample size determination

For example, if we are considering three potential dichotomous modifiers/confounders, we now have eight 2 by 2 tables.

There are many other scenarios still to come [more than one exposure, measured exposures, measured potential modifiers/confounders, additive but required variables]