Likelihood Ratios

A *likelihood ratio* is a ratio of two probabilities. These two probabilities have a particular form and we will soon see what that general form is.

Likelihood ratios are a very important in statistics and play a large role in many different theoretical arenas. In general we derive likelihood ratios from likelihood functions, but simple examples of likelihood ratios are fairly common.

Likelihood ratios are important because they provide a means of measuring the evidence in the data for one hypothesis over another.

To illustrate consider the following example:

This 2x2 table gives the properties of a diagnostic test for the presence of some disease.

		Test Result	
		positive	negative
Disease	yes	0.94	0.06
Status	no	0.02	0.98

The sensitivity of the test is P(T+|D+)=0.94 and The specificity is P(T-|D-)=0.98.

Now suppose this test is preformed on a person and the *test result is positive*. A physician might then ask:

1) Should this observation lead me to believe that this person has the disease?

2) Does this observation justify my acting as if he has the disease?

3) Is this test result evidence that he has the disease?

These generic questions define three distinct problem-areas of statistics:

1. What should I believe? (Bayesian Inference)

2. What should I do? (Decision Theory)

3. What do these data say? (Evidential Analysis)

Answering question #1 requires applying Bayes theorem, which we'll learn in the next lecture. Question #2 is answered with the Frequentist hypothesis testing. Likelihood ratios provide the mechanism for answering question #3.

For now, it is enough to understand that likelihood ratios tell us what the data say about one hypothesis versus another. That is, they provide the answer for question #3 (and hence not any other question).

Returning to our example…

		Test Result	
		positive	negative
Disease	yes	0.94	0.06
Status	no	0.02	0.98

After observing a *positive test result*, the physician's answers to the three questions must be:

1. Maybe
2. Maybe
3. Yes

Why?

The answers to questions #1 and #2 depend on more than just the test outcome itself. Hence the answer of 'maybe'.

Specifically, the answer to question #1 depends on what the physician believes prior to conducting the test (is the disease rare?) and the answer to question #2 depends on the risk-benefit tradeoff associated with the treatment (What are the side effects of treatment? What's the consequence of a type I error? … a type II error?).

Regardless of these extraneous considerations (risks, benefits, prior beliefs etc.), we are always correct when we interpret a positive result from this test as evidence that the disease is present.

Why? Our reasoning here is intuitive because of the context (for example no one would argue that a positive test is evidence that the disease is absent.)

But it is also intuitive in a statistical sense: if the disease was really present, the probability of observing a positive result is 0.94
($P(T+|D+)=0.94$)

and if the disease was really absent, the probability of observing a positive result is 0.02
($P(T+|D-)=1-P(T-|D-)=1-0.98=0.02$)

and, finally, 0.94 is greater than 0.02.

⇨ Thus we are more likely to observed a positive test result when the disease is present, and hence a positive result is evidence for the hypothesis that disease is present versus that the disease is absent.

This reasoning leads to a general principle:

The Law of Likelihood

If hypothesis A implies that the probability of observing some data X is $P(X|A)=P_A(X)$, while hypothesis B implies that the probability is $P(X|B)=P_B(X)$, then the observation X=k is **evidence supporting A over B if $P_A(k) > P_B(k)$**, and the likelihood ratio, $P_A(k)/P_B(k)$, measures the strength of that evidence.

- Likelihood ratios, $LR = P_A(k)/P_B(k)$, measure the strength of the evidence

- "H_A is supported over H_B by a factor of LR."

  a. If LR=1, the evidence for H_A vis-à-vis H_B is neutral

  b. If LR>1, the evidence supports H_A over H_B

  c. If LR<1, the evidence supports H_B over H_A

Now we see that a likelihood ratio is a ratio of two probabilities, where each probability gives the probability of the (observed) data under different hypotheses.

The degree to which the evidence (or data) supports one hypothesis over another is also important.

For interpreting and communicating the strength of evidence it is useful to divide the LR scale into descriptive categories (although we won't talk about how these benchmarks were derived).

The benchmarks are LRs of 8 and 32 .

- Weak evidence
  - for H_A over H_B: $1<LR<8$
  - for H_B over H_A: $1/8<LR<1$
- Moderate evidence
  - for H_A over H_B: $8<LR<32$
  - for H_B over H_A: $1/32<LR<1/8$
- Strong evidence
  - for H_A over H_B: $32<LR$
  - for H_B over H_A: $LR<1/32$

In our diagnostic example the likelihood ratio of interest is 47, indicating strong evidence.

		Test Result positive	negative
Disease	yes	0.94	0.06
Status	no	0.02	0.98

A positive test result is statistical evidence supporting $\mathbf{H_{D+}}$ over $\mathbf{H_{D-}}$ because

$$LR = \frac{P(T+ \mid H_{D+})}{P(T+ \mid H_{D-})} = \frac{0.94}{0.02} = 47$$

Hence the answer to question #3, "Is this test result evidence that he has the disease?", is correctly answered in the affirmative (and we now know why!).

Unfortunately, statistical evidence can be misleading.

For example, it is possible to observe a positive test when the disease is in fact absent.

In such a situation, we would still interpret the positive test as evidence that the disease is present (because we don't know the true disease status).

This is ok, and it is important to understand that our interpretation of the evidence is *correct regardless of the true disease status*. It is the evidence itself that is misleading. We have not made an 'error' in how we interpreted the evidence.

How likely a likelihood ratio is to be misleading depends on how strong the evidence is. It is much less likely that strong evidence is misleading than weak evidence is.

Here, this diagnostic test is a good one in the sense that misleading evidence is seldom observed because $P(T+|D-)=0.02$.

Just for fun, consider a different diagnostic test for the same disease (call it test #2). Its properties are listed in the table below.

Test #2		Test Result	
		positive	negative
Disease	yes	0.47	0.53
Status	no	0.01	0.99

A positive result on the second test is again statistical evidence supporting $\mathbf{H_{D+}}$ over $\mathbf{H_{D-}}$ by a factor of 47:

$$LR = \frac{P(T+|H_{D+})}{P(T+|H_{D-})} = \frac{0.47}{0.01} = 47$$

But the probability of observing misleading evidence under this second test is half of that of the first test because P(T+|D-)=0.01.

This leads to a natural and very important **question**:

Is a positive result on the second test stronger evidence in favor of disease than a positive result on the first one?

Or

Is the positive result on the second test “less likely to be misleading”, “more reliable” in some sense, or does it warrant more “confidence”?

Answer:

No! A positive result on the second test is equivalent, as evidence about the presence or absence of disease, to a positive result on the first.

(For those disbelievers we next prove it with a simple application of Bayes Theorem.)

Proof:

A positive test result is misleading *if and only if* the subject does not have the disease. The probability of that P(D-|T+).
Claim: P(D-|T+) is the same for both tests!

Applying Bayes Theorem:

$$P(D-|T+)=\frac{P(T+|D-)P(D-)}{P(T+|D-)P(D-)+P(T+|D+)P(D+)}$$

$$=\frac{P(D-)}{P(D-)+\dfrac{P(T+|D+)}{P(T+|D-)}P(D+)}=\frac{P(D-)}{P(D-)+47\cdot P(D+)}$$

where P(D+) is the prevalence of the disease. So we see that although P(D-|T+) depends on the prevalence, it is the same for both tests because the strength of evidence (i.e., likelihood ratio) is the same in both cases.

⇨ Thus, an observed positive result is no more likely to be misleading if it comes from one test than if it comes from the other.

This is, in fact, old news (although we tend to ignore it everyday when we interpret statistical results). R.A. Fisher, one of the founding fathers of modern statistics and ardent proponent of the p-value, wrote:

"In fact, as a matter of principle, the infrequency with which, in particular circumstances, decisive evidence is obtained, should not be confused with the force, or cogency, of such evidence."

-- R. A. Fisher, 1959, p.93

Up to now, we have considered only the case when there were just two hypotheses of interest.

What happens if we want to characterize the evidence about, say a probability, a rate or a mean?

In this case there are an infinite number of likelihood ratios because the parameter of interest may take an infinite number of values.

To deal with this situation we need an additional concept: the likelihood function.

Likelihood Functions

A likelihood function is (essentially) a function that gives the probability of our data under a specified hypothesis.

Example: Suppose that I was interested in measuring the evidence about the probability of cardiovascular death within one year of taking drug A.

Suppose further that in a Cincinnati clinic I gave the drug to 22 people and 7 died within the following year.

I am interested in learning about the probability of cardiovascular death within one year of taking drug A. Let's call that unknown probability θ. I might want to know if that probability is 20% or 30%.

Now, if the probability of cardiovascular death on this drug is 20% ($\theta = 0.2$), the chance of observing exactly 7 CV deaths out of 22 people is given by our favorite binomial formula

$$P(Y = 7|\theta = 0.2) = \binom{22}{7}(0.2)^7(0.8)^{15} = 0.0768$$

(This assumes, of course, that the binomial model's assumptions are met. Quick quiz: the required constant probability of success and independent trials translates into what here?)

And if the probability of cardiovascular death on this drug is 30% ($\theta = 0.3$), the chance of observing exactly 7 CV deaths out of 22 people is given by

$$P(Y = 7|\theta = 0.3) = \binom{22}{7}(0.3)^7(0.7)^{15} = 0.1771$$

(Notice that I've used the conditional probability sign '|' to emphasize the fact that I am assuming I know θ to be some particular value. That is, I'm conditioning on θ.)

The Law of Likelihood says that we measure the evidence for the hypothesis that the probability of CV death is 30% versus 20% with the ratio 0.1771/0.0768 = 2.31.

Written differently, we have that observing 7 CV deaths out of 22 people is evidence supporting the hypothesis that θ = 0.3 over the hypothesis that θ = 0.2 by a factor 2.31 (weak evidence).

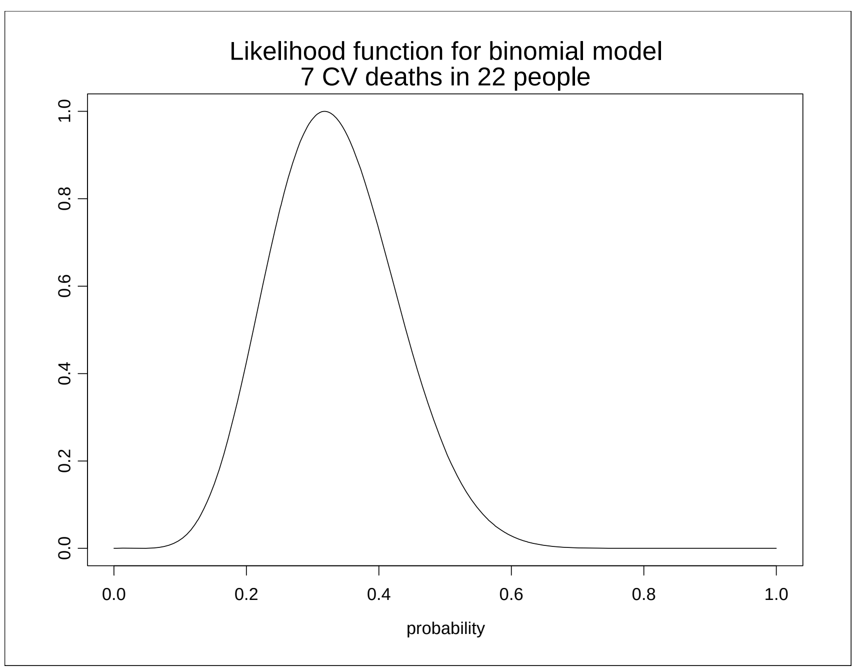
$$
\begin{aligned}
LR \overset{def}{=} \frac{L(0.3)}{L(0.2)} \overset{def}{=} \frac{L(0.3 \mid Y = 7)}{L(0.2 \mid Y = 7)}
&= \frac{\mathrm{P}(Y = 7 \mid \theta = 0.3)}{\mathrm{P}(Y = 7 \mid \theta = 0.2)} \\
&= \frac{\binom{22}{7} 0.3^{7} (0.7)^{15}}{\binom{22}{7} 0.2^{7} (0.8)^{15}} \\
&= 2.31
\end{aligned}
$$

The **likelihood function** is simply

$$L(\theta) = L(\theta|Y=7) = P(Y = 7|\ \theta\ ) = \binom{22}{7}\theta^7(1-\theta)^{15}$$

Notice the change in notation from 'P' to 'L' in an attempt to emphasize that (1) the data are now observed, (2) θ is now a variable and (3) the likelihood function is no longer a true probability function.

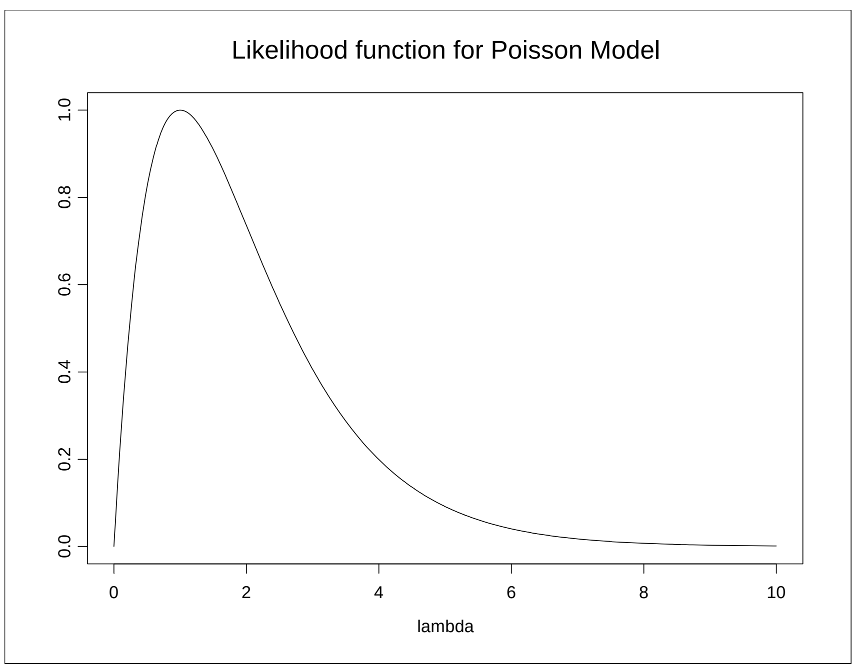
Rather than listing all the likelihood ratios (which would be quite cumbersome) we simply plot the likelihood function, L(θ|Y=7), as a function of θ. We'll standardized the y-axis so the peak is at 1.



If we are interested in learning about a rate (say after observing K events in a specified time period), then the Poisson distribution gives the appropriate likelihood function as

$$L(\lambda) = L(\lambda|Y=k) = P(\ Y = k\ |\lambda) = \lambda^{k}e^{-\lambda}/k!$$

For example, suppose I purchase a bunch of lotto tickets each month and got one winner. I am interested in the rate of winning lotto tickets for my buying habits. Then $L(\lambda|Y=1)$ is plotted below.

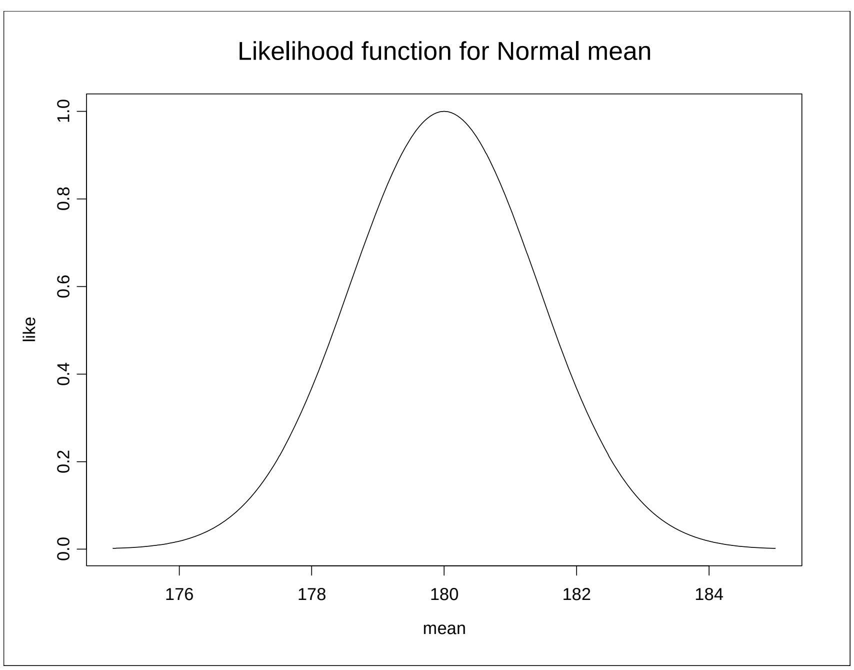


And if we are interested in learning about the mean of a distribution (say after observing a score of k=180, with fixed $\sigma^2=2$), then the Normal distribution gives the appropriate likelihood function as

$$L(\mu|\sigma^2) = L(\mu|\sigma^2, Y=k) = P(Y = k|\mu,\sigma^2)$$

$$L(\mu|\sigma^2, Y=k) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-\mu}{\sigma}\right)^2}$$

The likelihood function is then



When we talk about likelihood functions we usually refer to two quantities: (1) the value of the parameter that gives the maximum probability (called the Maximum Likelihood Estimator - MLE) and (2) the curvature or peakedness of the likelihood function (called the information).

So far we have considered only very simple likelihood functions (based on only one observation). In the future we will encounter likelihood functions for groups of observations.

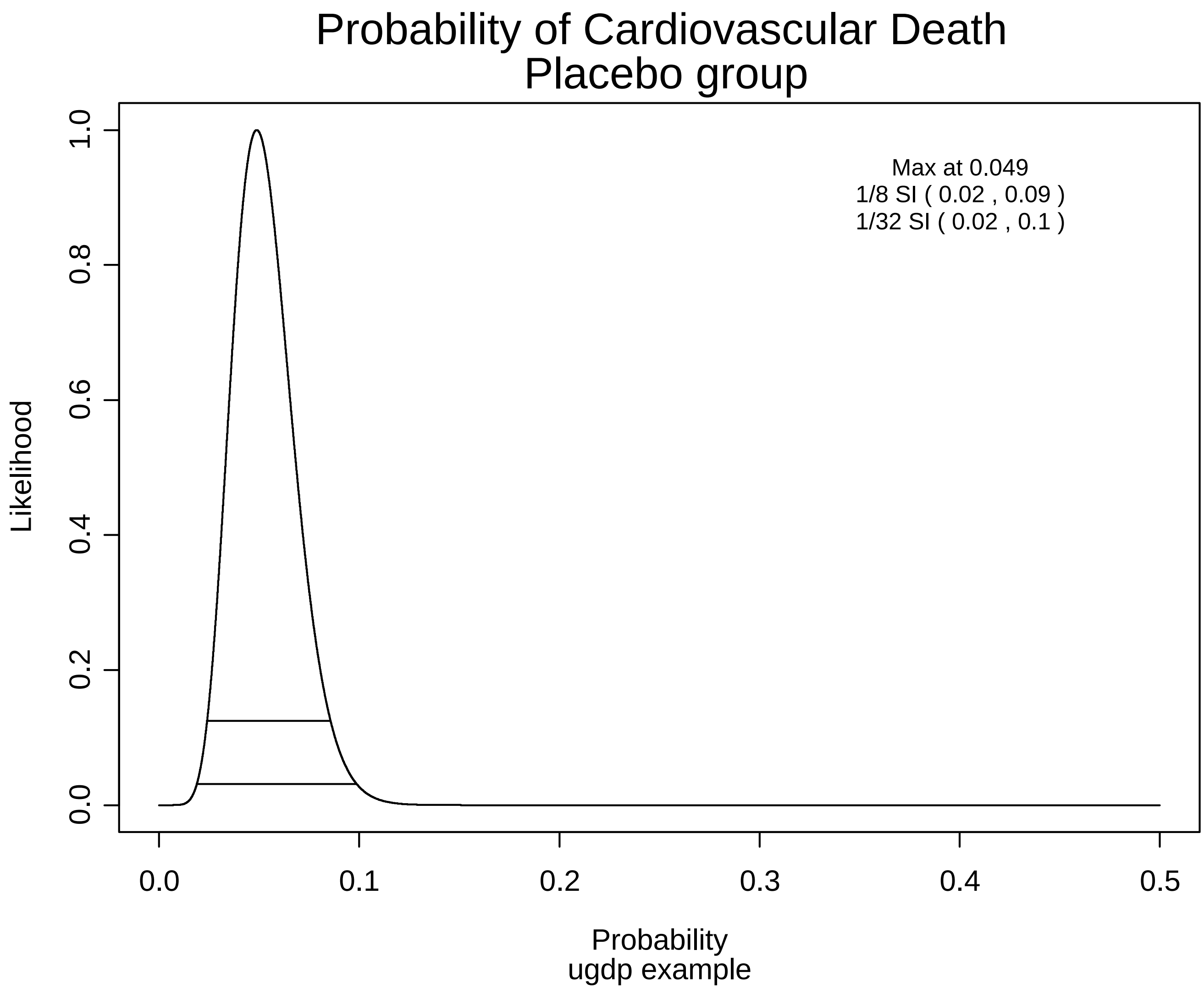
Finally, note that all probability distribution functions are likelihood functions if you consider the data fixed and the parameters as variables.

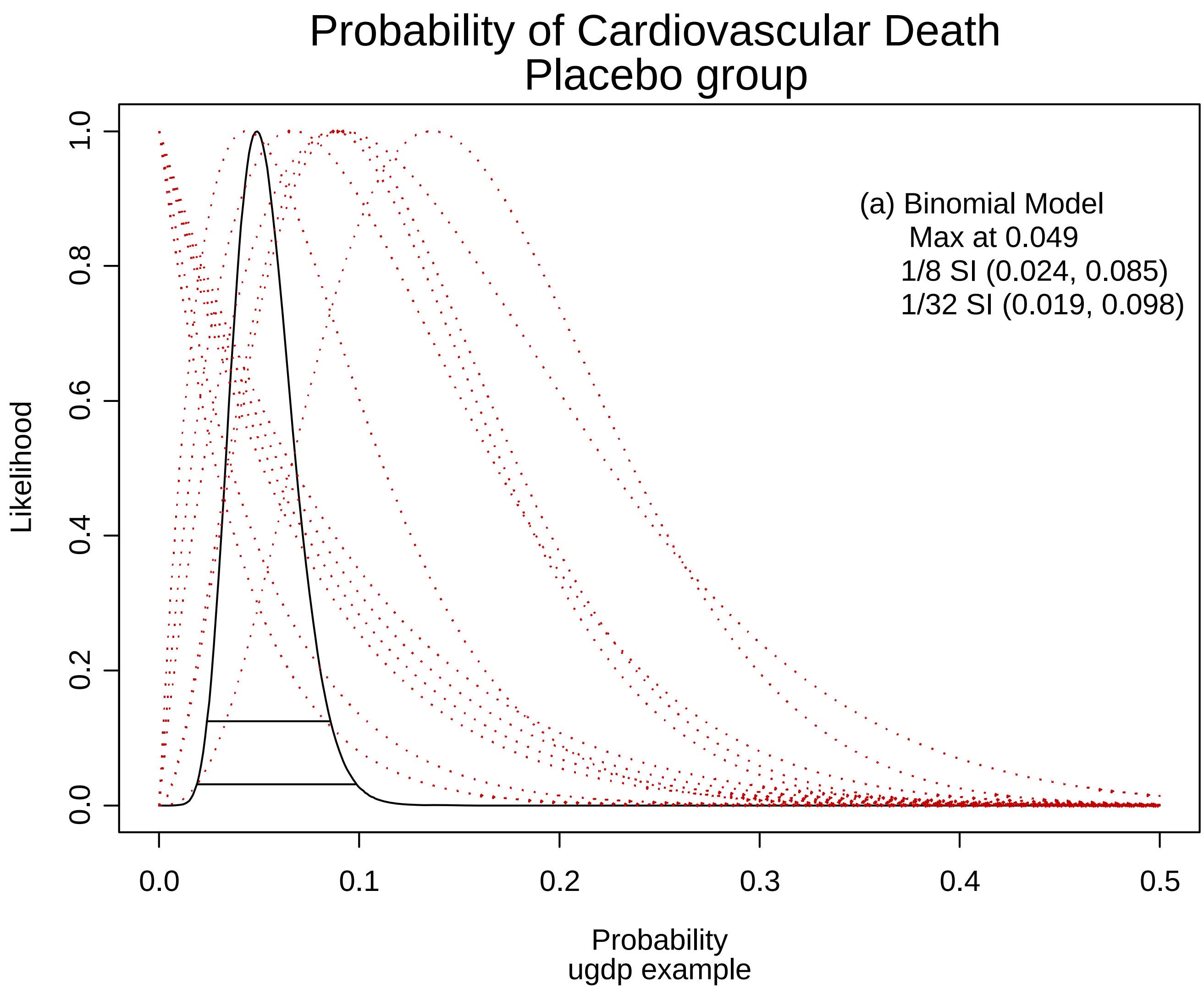
The University Group Diabetes Program (1961-1975)

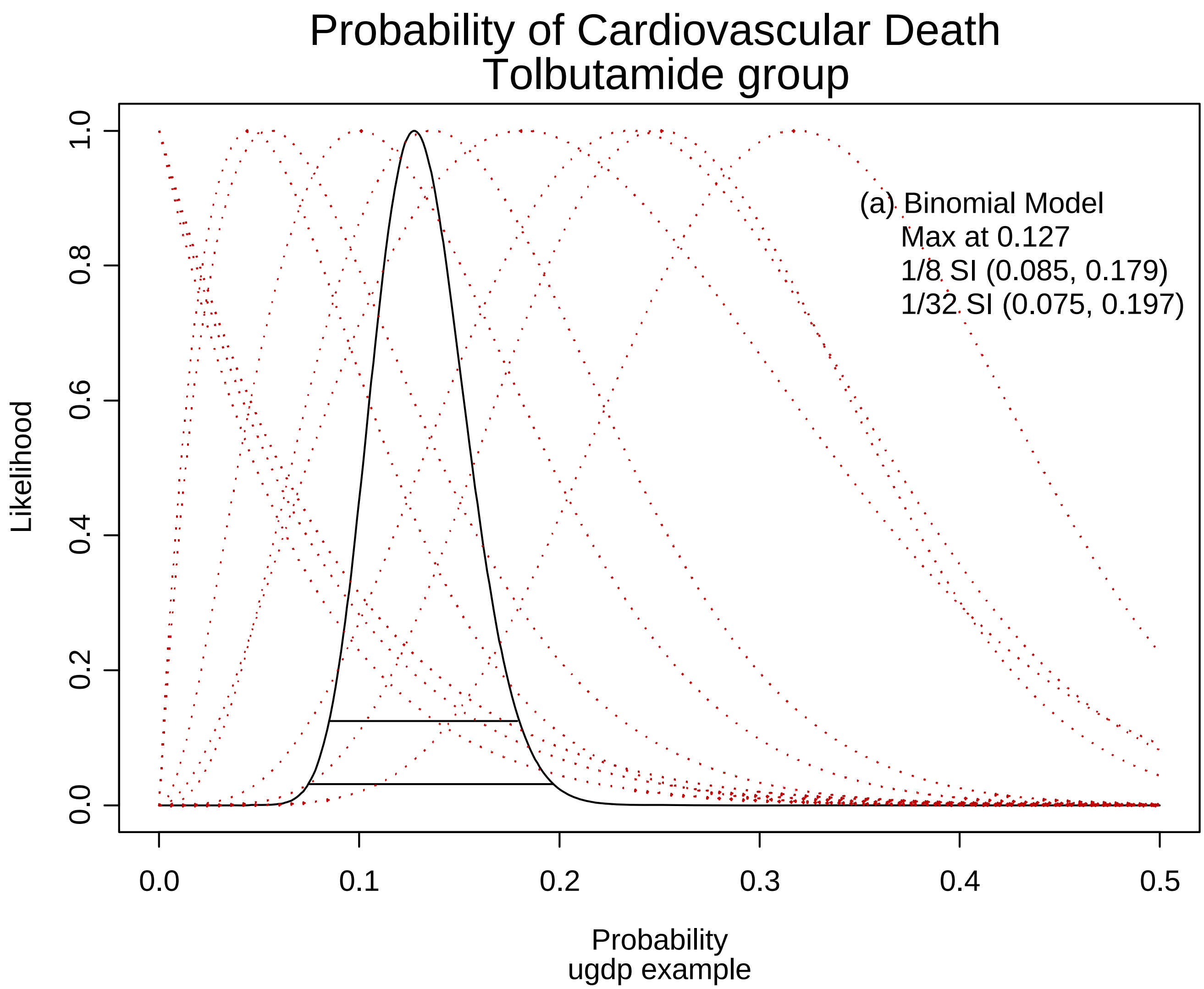
- Multi-centered, randomized clinical trial, to evaluate the effect of Tolbutamide on vascular complications of adult-onset diabetes.

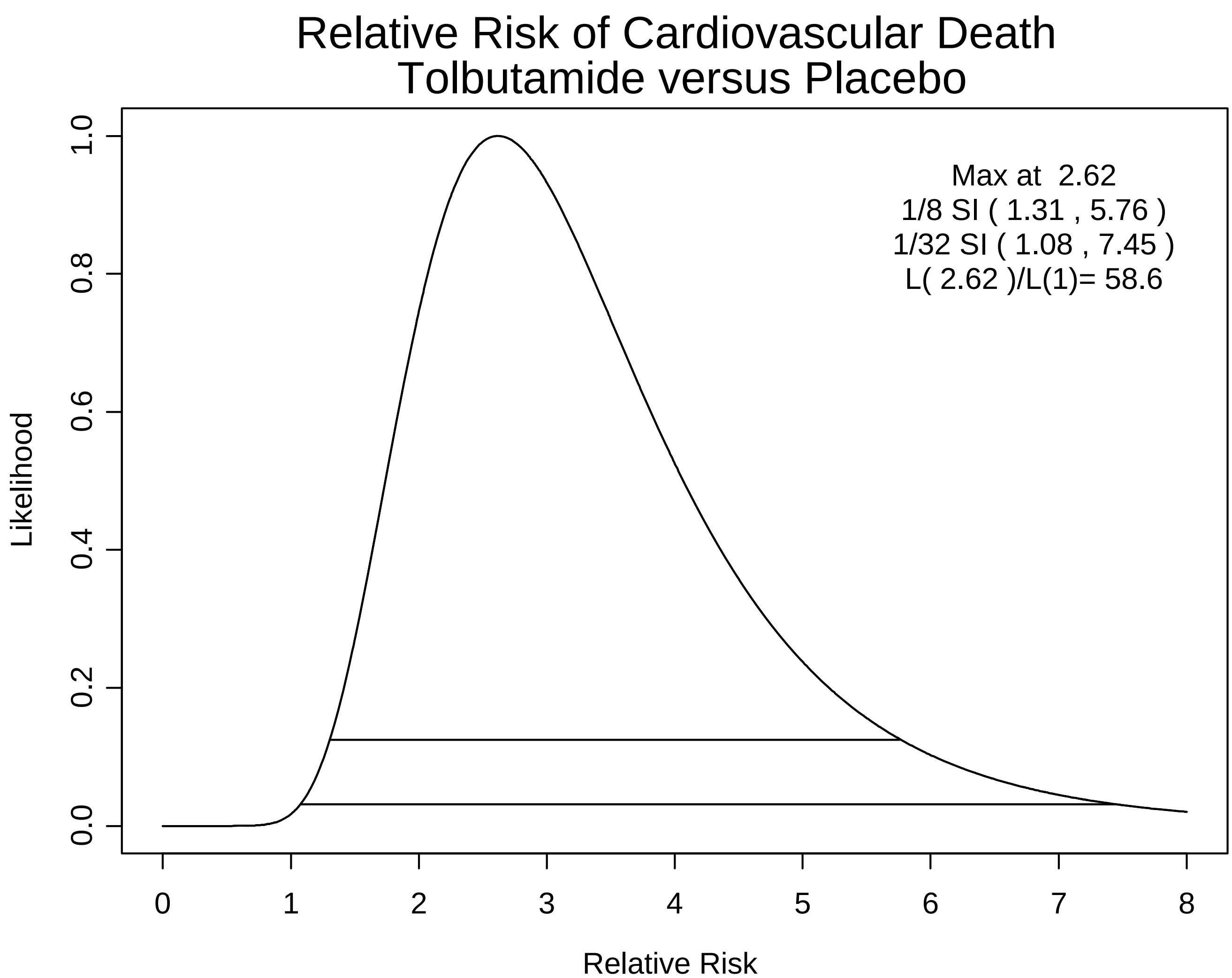
- Probability of cardiovascular death?

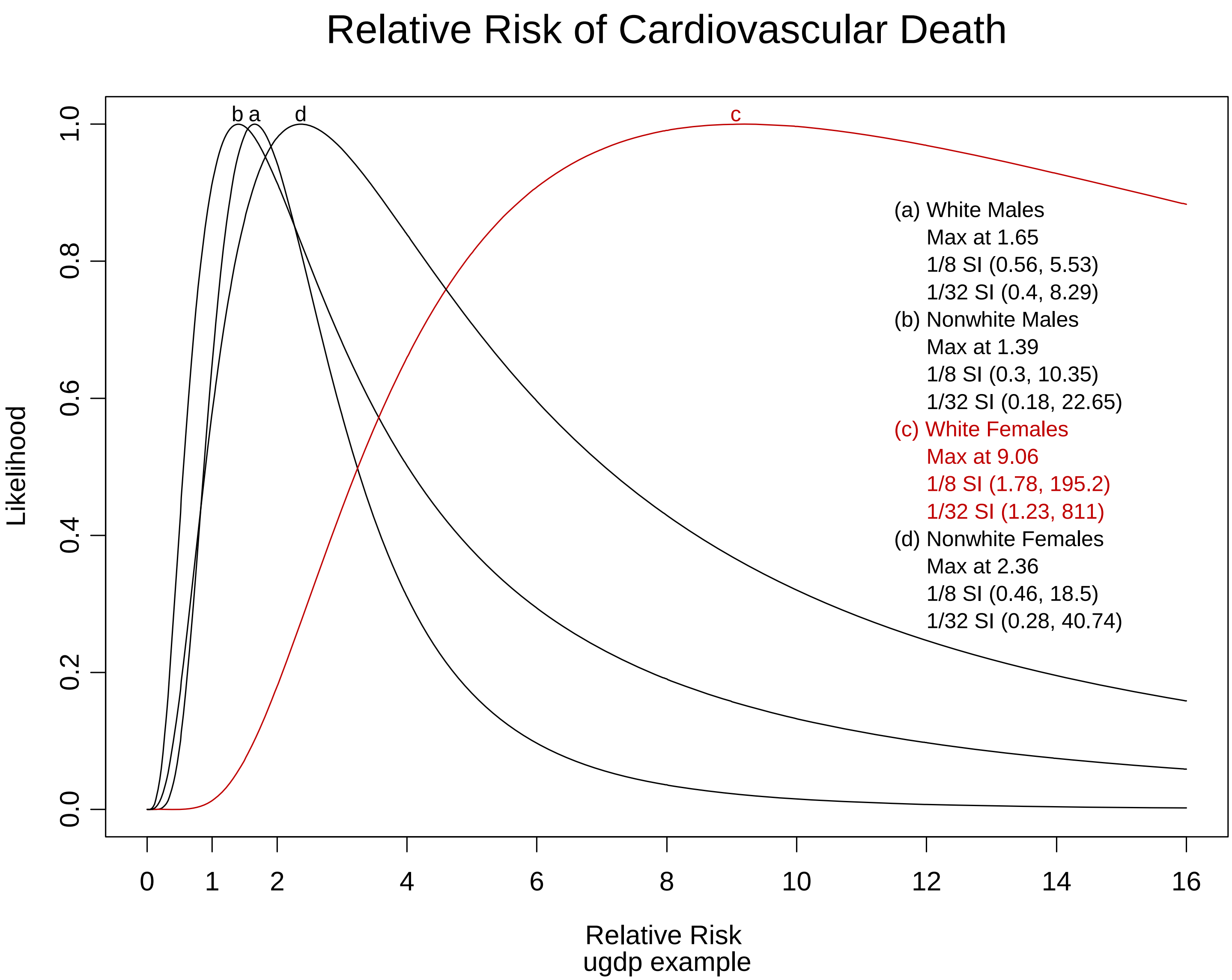
Center	Tolbutamide		Placebo	
	Deaths	Total	Deaths	Total
Baltimore	1	22	0	24
Cincinnati	7	22	2	23
Cleveland	1	18	0	19
Minneapolis	6	24	2	22
New York	2	20	3	22
Williamson	3	22	1	23
Birmingham	2	11	0	12
Boston	4	17	1	15
Chicago	0	12	1	11
St. Louis	0	11	0	10
San Juan	0	12	0	12
Seattle	0	11	0	11
All Centers	26	204	10	205


Probability of Cardiovascular Death
Placebo group
Max at 0.049
1/8 SI (0.02 , 0.09)
1/32 SI (0.02 , 0.1)
Likelihood
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Probability
ugdp example


Probability of Cardiovascular Death
Placebo group
(a) Binomial Model
Max at 0.049
1/8 SI (0.024, 0.085)
1/32 SI (0.019, 0.098)
Likelihood
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Probability
ugdp example


Probability of Cardiovascular Death
Tolbutamide group
Likelihood
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Probability
ugdp example
(a) Binomial Model
Max at 0.127
1/8 SI (0.085, 0.179)
1/32 SI (0.075, 0.197)


Relative Risk of Cardiovascular Death
Tolbutamide versus Placebo
Max at 2.62
1/8 SI (1.31 , 5.76)
1/32 SI (1.08 , 7.45)
L(2.62)/L(1)= 58.6
Likelihood
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
7
8
Relative Risk


Relative Risk of Cardiovascular Death
Likelihood
0.0
0.2
0.4
0.6
0.8
1.0
b a d c
(a) White Males
Max at 1.65
1/8 SI (0.56, 5.53)
1/32 SI (0.4, 8.29)
(b) Nonwhite Males
Max at 1.39
1/8 SI (0.3, 10.35)
1/32 SI (0.18, 22.65)
(c) White Females
Max at 9.06
1/8 SI (1.78, 195.2)
1/32 SI (1.23, 811)
(d) Nonwhite Females
Max at 2.36
1/8 SI (0.46, 18.5)
1/32 SI (0.28, 40.74)
0 1 2 4 6 8 10 12 14 16
Relative Risk
ugdp example