Models In Epidemiology And Biostatistics
Gordon Hilton Fick
Session 1: Preliminaries to Models

Important choices at the design stage of a project

Success or Failure: Which Rate? Why?

Probability or Odds: Which one? Why?

Difference or Ratio: Which one? Either? Why?

Modification or Confounding: The first steps

Interpretation: Impact of the choices

Careful review of the actual tables before reporting measures of association.

Impact of the measure

Counterfactuals

An example to provide context

A questionnaire was given to patients after surgery. The instrument was developed to measure post-operative outcomes.

Let us suppose that the investigators constructed a measure of ‘success’ from the questionnaire.

Further, all patients received 1 of 2 different pre-operative treatments.

Notation

Intervention: pre-op treatment T : T_1 or T_2

Outcome: success: S or F

Population characteristic: the probability of success: P(S)

This probability may depend on several things.

Maybe the probability depends on T:

We then consider conditional probabilities:

The probability of success GIVEN the treatment: $P(S \mid T_1)$

may not equal $P(S \mid T_2)$

Let $\mathrm{p}_1 = P(S \mid T_1)$ and let $\mathrm{p}_2 = P(S \mid T_2)$

Risks

Probabilities are sometimes called risks. We might speak of 'risk' when the outcome is something negative like mortality.

For example, the 'probability of death' is often called the 'risk of death'

We would not speak of the 'risk of survival' though.

Rates

Probabilities are sometimes called rates

We then speak of success rates, failure rates, mortality rates, remission rates

The outcome could be positive or negative

Differences

A rate difference: RD = $P(S \mid T_1)$ - $P(S \mid T_2)$
so that $P(S \mid T_1)$ = RD + $P(S \mid T_2)$
The success rate for T_1 is RD more than T_2.
For example: Say that RD=.2. Then we could say that: the success rate for T_1 is 0.2 more than T_2

Ratios

A rate ratio: $RR = P(S \mid T_1) / P(S \mid T_2)$

so that $P(S \mid T_1) = RR * P(S \mid T_2)$

The success rate for T_1 is RR times the success rate for T_2 .

For example: Say that RR=1.4. The we could say that: the success rate for T_1 is 1.4 times the success rate for T_2

Or that: the success rate for T_1 is 140% of the success rate for T_2

It would be incorrect to say that: the success rate for T_1 is 1.4 more than the success rate for T_2. This statement implies a 'difference' between rates, not a ratio of rates.

Statements with percentages. Beware!

To continue the last comment:

It is (at least) confusing and often misinterpreted if one says that:

The success rate for T_1 is 40% more than the success rate for T_2 .

(40% in relative terms? 40% in absolute terms?)

If A=RR*B, then A-B = (RR-1)*B which is not the same as A-B = (RR-1)

Risk

Now we can also (and often do) think in terms of failure rates:

Risk difference: $RD = P(F \mid T_1) - P(F \mid T_2)$

so $P(F \mid T_1) = RD + P(F \mid T_2)$

Risk Ratio: $RR = P(F \mid T_1) / P(F \mid T_2)$

so $P(F \mid T_1) = RR * P(F \mid T_2)$

Ratios of Probabilities, Rates, Risk and Health

Maybe the best term is “Probability Ratio” but this term is rarely seen in the health research literature

When it is clear that the outcome is the negative one [for example: failure, disease, recurrence], then risk is a synonym for probability. So we have risk ratios

We should not speak of the 'risk of success' or the 'risk of health'. Sometimes one sees the term 'health ratio' [for example: Stata uses health ratio]

At times, one sees 'rate ratio' when it is clear one means a ratio of probabilities.

Odds

Odds = $\dfrac{p}{1-p}$

Conditional Odds = $\dfrac{P(S \mid T_i)}{1-P(S \mid T_i)}$

Odds of disease

Odds of exposure

Odds is not a probability

Odds can be any positive number.

For example, if $p = 2/3$, then odds = 2

If $p = 4/5$, then odds = 4

Odds and probability can be 'very' different.

Odds can be 'close' to a probability

For example, if $p=1/20$, then odds $=1/19$

If $p=1/100$, then odds $= 1/99$

In fact, if p is 'small', then 1-p is 'close' to 1

So that odds is 'close' to p

In spite of this 'closeness', it is best to keep the 2 terms separate

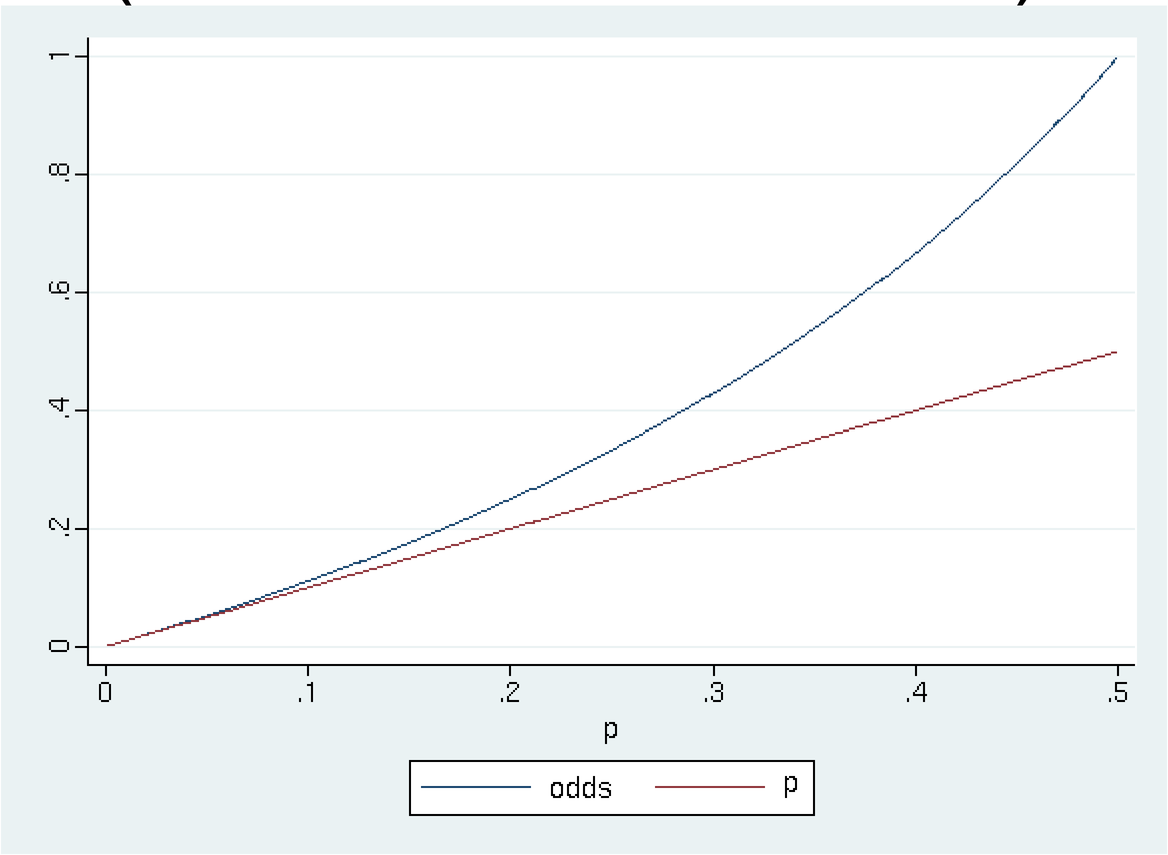
If you mean odds, say so

If you mean probability, say so

If you say risk (or chance), you mean probability (not odds)

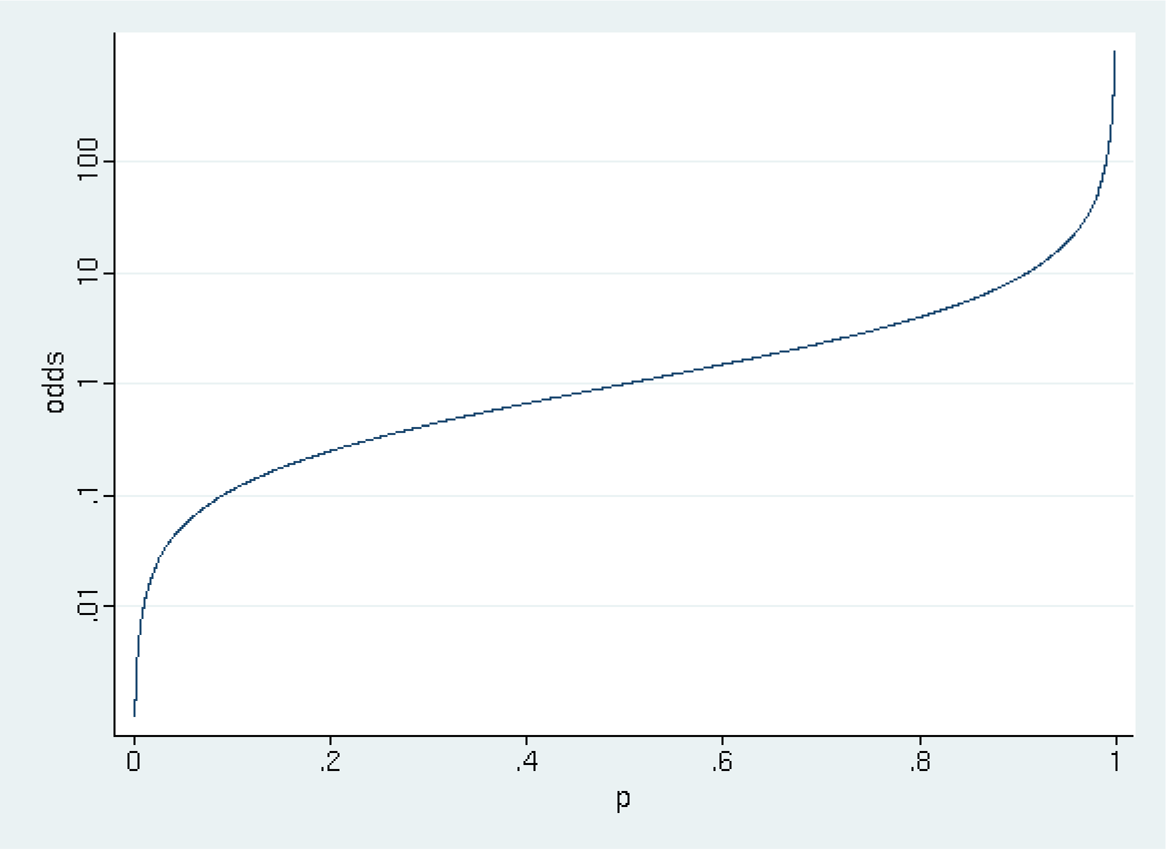
Graphs of Odds versus P

(when P is less than 0.5)



Graph of Odds versus P

(notice logarithmic scale on 'Odds' axis)



Odds Ratio

For example: $$\mathrm{OR} = \frac{\mathrm{Odds}(S \mid T_1)}{\mathrm{Odds}(S \mid T_2)}$$

So that the odds of success with T_1 is OR times the odds of success with T_2

For example: Say OR=3. Then, we could say that the odds of success with T_1 is 3 times the odds of success with T_2

Do not say 'risk, likely or probability' if you mean 'odds'

The 'Magic' of Logarithms: Ratios to Differences

$$\log(\mathrm{OR}) = \log \mathrm{odds}(S \mid T_1) - \log \mathrm{odds}(S \mid T_2)$$

The logarithm of an odds ratio is the difference between two log odds. A log odds is called a logit.

$$\log(\mathrm{RR}) = \log P(S \mid T_1) - \log P(S \mid T_2)$$

The logarithm of a rate ratio is the difference between two log rates.

If a ratio = 1, then the log of the ratio = 0

If a ratio is less than 1, then the log of the ratio is negative

If a ratio is greater than 1, then the log of the ratio is positive

…and so a study was conducted

2200 patients completed the questionnaire: 1100 received T_1 and 1100 received T_2

Here are some results:

	T_1	T_2
Success	**195**	**505**
Failure	**905**	**595**

Adapted from J.G.Kalbfleisch Probability and Statistical Inference, Volume 2 Springer-Verlag NY (pg.180-181)

… from this table, we can compute

…estimates of the probabilities discussed earlier:

$P(S \mid T_1)$ and $P(S \mid T_2)$

$$\hat{p}_1 = \hat{P}(S|T_1) = \frac{195}{1100} \approx 0.1773$$

$$\hat{p}_2 = \hat{P}(S|T_2) = \frac{505}{1100} \approx 0.4591$$

Inference?

We can see that $\hat{p}_1 \neq \hat{p}_2$

But does this mean that $p_1 \neq p_2$?

This brings us to the notions of error and bias

Notice: the “hats” are important.

Confusion reigns if the same symbol is used for the sample characteristic and the population characteristic.

Sometimes Greek letters are used for the population characteristic and the Latin equivalent is used for the estimate.

Let us consider error first

We would like to know the probabilities.

But, except in simple situations like card games and casino games, we can never ‘know’ the probabilities. They are always behind the curtain. They will always be ‘unknowns’.

They are unknowns partly because of sampling error.

Sampling error

2200 patients is a lot of patients but:

1) there will be more patients

2) patients do not respond the same way

3) the probabilities are about certain 'populations' of patients

4) we must always distinguish between populations and 'samples' from these populations

5) samples provide us with estimates of population characteristics (probabilities)

Estimates of RD and RR

We have an estimate of RD:

$$\widehat{RD} = \hat{p}_1 - \hat{p}_2 = 0.1773 - 0.4591 = -0.2818$$

and we have an estimate of RR:

$$\widehat{RR} = \hat{p}_1 / \hat{p}_2 = 0.1773 / 0.4591 = 0.3861$$

Estimate of Odds Ratio

Estimated Odds of success with T_1 =

$$\frac{195/1100}{905/1100} = \frac{195}{905} \approx 0.2155$$

Estimated Odds of success with T_2 =

$$\frac{505/1100}{595/1100} = \frac{505}{595} \approx 0.8487$$

$$\hat{\mathrm{OR}} = \frac{\frac{195}{1100}}{\frac{905}{1100}} / \frac{\frac{505}{1100}}{\frac{595}{1100}} = \frac{195 * 595}{905 * 505} \approx 0.2539$$

Bias

In health research, sampling error issues are often minimal compared with bias issues.

Arguably, the most serious form of bias is one that goes by many names. The oldest name might be Simpson's paradox.

Let us return to the surgery example.

Surgery type

In fact, the data described previously came from a study where the surgeons were performing 2 quite different types of surgery: Surgery 1 (A_1) and Surgery 2 (A_2).

Here is the data now:

	Surgery			
	A1		A2	
	Treatment		Treatment	
	T1	T2	T1	T2
Success	100	5	95	500
Failure	900	95	5	500

The Estimates

	A1		A2	
	Treatment		Treatment	
	T1	T2	T1	T2
Success	100	5	95	500
Failure	900	95	5	500
Rate Estimates	0.1	0.05	0.95	0.50
Rate Differences	0.05		0.45	
Rate Ratio	2		1.9	
Odds Estimates	1/9	1/19	19	1
Odds Ratio	2.1111		19.0	

The earlier ‘simpler’ analysis

	Treatment	
	T1	T2
Success	195	505
Failure	905	595
Estimate	0.1773	0.4591
Rate Difference	-0.2818	
Rate Ratio	0.3861	
Odds Ratio	0.2539	

What is going on?

The 2 stratum specific (surgery specific) rate ratios are nearly the same (2 cf. 1.9) BUT very different from the crude (‘simpler’) rate ratio (0.39).

The stratum specific rate differences are not even close to one another. (0.05 cf. 0.45)

The stratum specific odds ratios are very different. (2.1 cf. 19)

Here is the arithmetic

$$\frac{195}{1100} = \frac{100}{1100} + \frac{95}{1100} = \frac{100}{1000}\frac{1000}{1100} + \frac{95}{100}\frac{100}{1100}$$

$$\frac{505}{1100} = \frac{5}{1100} + \frac{500}{1100} = \frac{5}{100}\frac{100}{1100} + \frac{500}{1000}\frac{1000}{1100}$$

	A1	A2	Total
T1	1000	100	1100
T2	100	1000	1100

The population versions

Each row is using:

$$P(S)=P(SA_1)+P(SA_2)=P(S \mid A_1)P(A_1)+P(S \mid A_2)P(A_2)$$

For the first row on the previous page, for those receiving T1 (given T1)

Then for the second row, for those receiving T2 (given T2)

Given pretreatment T_1:

$$P(S \mid T_1) = P(SA_1 \mid T_1) + P(SA_2 \mid T_1)$$
$$= P(S \mid A_1 T_1) P(A_1 \mid T_1) + P(S \mid A_2 T_1) P(A_2 \mid T_1)$$

so for the data in this study, we get:

$$\frac{195}{1100} = \frac{100}{1100} + \frac{95}{1100}$$
$$= \frac{100}{1000}\frac{1000}{1100} + \frac{95}{100}\frac{100}{1100}$$

: a weighted sum of 100/1000 and 95/100

: 100/1000 gets most of the 'weight' (1000/1100)

Given pretreatment T_2:

$$P(S \mid T_2) = P(SA_1 \mid T_2) + P(SA_2 \mid T_2)$$
$$= P(S \mid A_1 T_2) P(A_1 \mid T_2) + P(S \mid A_2 T_2) P(A_2 \mid T_2)$$

and for the data in this study, we get:

$$\frac{505}{1100} = \frac{5}{1100} + \frac{500}{1100}$$
$$= \frac{5}{100}\frac{100}{1100} + \frac{500}{1000}\frac{1000}{1100}$$

: a weighted sum of 5/100 and 500/1000

: 500/1000 gets most of the 'weight' (1000/1100)

The fractions

So, for example, the fraction 1000 / 1100 is determined by the circumstance that the surgeons in this study "prefer" to assign pretreatment 1 to patients about to receive surgery 1.
Maybe this "preference" is largely dictated by policy in a region/hospital and such policy may not be the norm in other jurisdictions.
We see such fractions as specific to this study so such proportions are not generalizable, per se, to other surgeons/jurisdictions.

For this study and maybe ONLY this study, we see that

1000 out of the 1100 patients [receiving surgery 1] had received pretreatment 1.

1000 out of the 1100 patients [receiving surgery 2] had received pretreatment 2.

But these two fractions might be quite different in other studies.

Probabilities and estimates of probabilities

The four 'surgery specific' numbers [100/1000, 95/100, 5/100, 500/1000] can be seen as estimates of four probabilities.

We see that the two numbers not specific to surgery [195/1100, 505/1100] are based on the surgery specific numbers but depend on the fractions.

How do we assess pretreatment?

Suppose the four surgery specific numbers do generalize to other circumstances but the fractions [the weights] do not extend to other circumstances.

Let us use the four surgery specific numbers but think of the two weights as unknowns.

So we could think of two lines

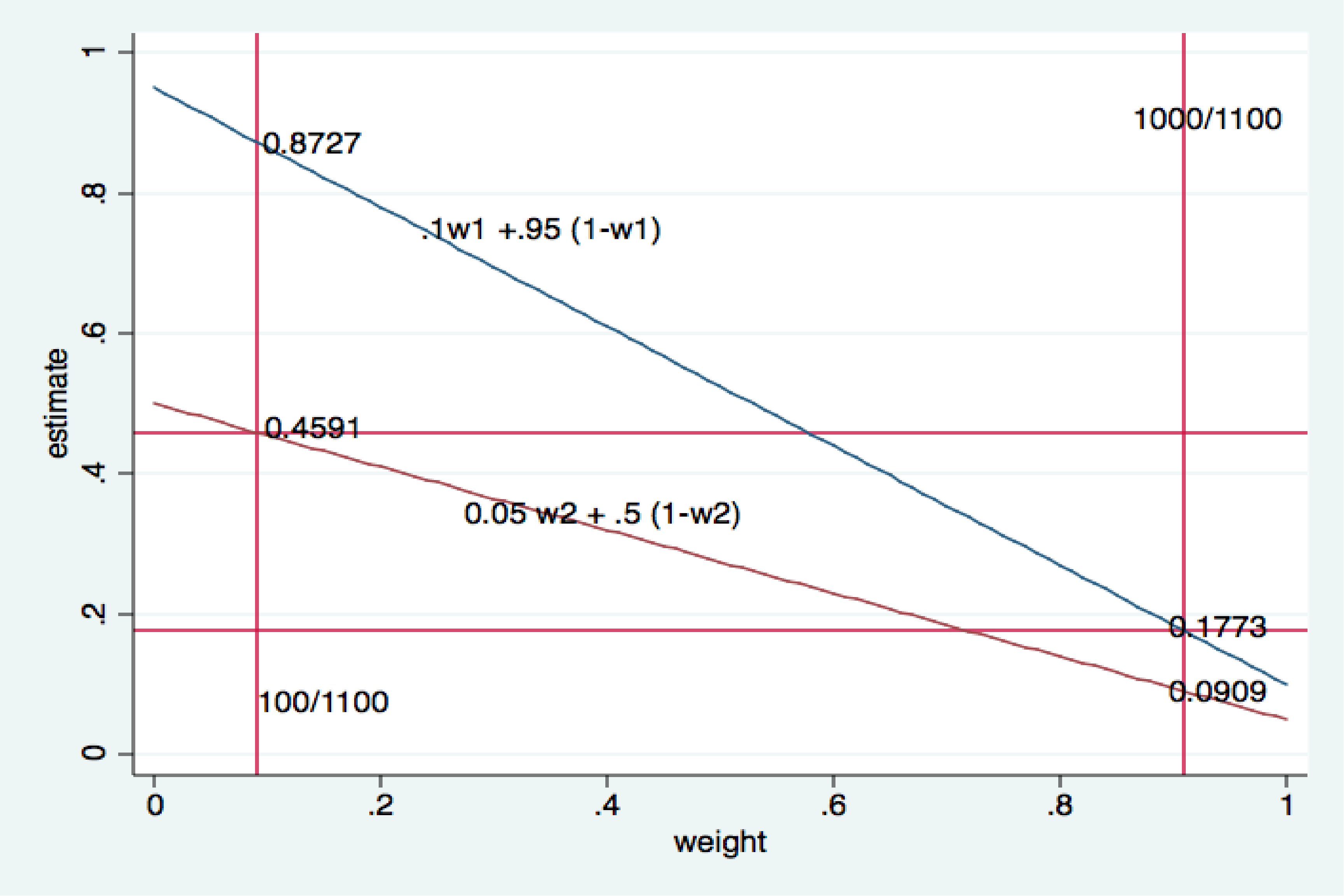
A blue line for those receiving T_1 :

$$\hat{p}_1(w_1) = w_1 \frac{100}{1000} + (1 - w_1) \frac{95}{100}$$

A red line for those receiving T_2 :

$$\hat{p}_2(w_2) = w_2 \frac{5}{100} + (1 - w_2) \frac{500}{1000}$$

A Comparison of Pretreatments [Blue versus Red]



Counterfactuals

It is sometimes argued that one should compare the 2 weighted sums at certain specified common values for the weights.

The weights 'observed' with this study are then 'real' while other weights provide for 'counterfactual'[2] considerations.

There are many possible comparisons and arguments for such comparisons.

[2] expressing what has not happened but could, would, or might under differing conditions

Counterfactual Plots

The two lines each contain three 'real' values

: at the endpoints when w = 0 for A_2 and 1 for A_1

: at the observed weight $w_1 = 0.9090 \text{ and } w_2 = 0.0909$

All the other points on these lines are 'counterfactual'.

In principle, one might consider the vertical distance between the two lines at any common weight.

Such a consideration is questionable as differences between the endpoints are different.

Risk of Failure or Probability of Success?

- You must decide how to view your outcome.
- Make the choice early on in your planning.
- Usually, you go with the choice made from your literature.
- Does it matter?
- Failure Rate Difference = - Success Rate Difference
- Odds (of Failure) Ratio = $\dfrac{1}{\text{Odds (of Success) Ratio}}$

  The Risk Ratio (aka Failure Rate Ratio) is not a function of the Success Rate ratio (...boooo!)

....and some of the (yay!)algebra ---->

Success or Failure?

Let the success probabilities be $\mathrm{S}_1 = P(S \mid T_1)$ and $\mathrm{S}_2 = P(S \mid T_2)$

so that the failure probabilities are $F_1 = 1 - S_1$ and $F_2 = 1 - S_2$

<u>For the Rate Difference:</u>

In terms of success, $\mathrm{RD(S)} = S_1 - S_2$

In terms of failure, $\mathrm{RD(F)} = F_1 - F_2$

$$(1 - S_1) - (1 - S_2) = S_2 - S_1 = -(S_1 - S_2)$$

<u>For the Odds Ratio:</u>

$$\mathrm{OR(F)} = \frac{F_1}{S_1} / \frac{F_2}{S_2} = \frac{F_1 S_2}{S_1 F_2} = \frac{1}{\dfrac{S_1}{F_1} / \dfrac{S_2}{F_2}} = \frac{1}{\mathrm{OR(S)}}$$

Success or Failure?

Alas, for the Rate Ratio:

$$\mathrm{RR(F)} = \frac{F_1}{F_2} = \frac{1-S_1}{1-S_2} = \frac{1-S_1}{S_1}\,\frac{S_2}{1-S_2}\,\frac{S_1}{S_2} = \frac{F_1}{S_1}\,\frac{S_2}{F_2}\,\frac{S_1}{S_2} = \mathrm{OR(F)}\ \mathrm{RR(S)}$$

$$\text{so that } \mathrm{RR(F)} = \frac{\mathrm{RR(S)}}{\mathrm{OR(S)}}$$

...aack! The rate ratio (in terms of failure) is the odds ratio (in terms of failure) times the rate ratio (in terms of success)

..or the rate ratio (in terms of failure) is the rate ratio (in terms of success) divided by the odds ratio (in terms of success)

The 2 rate ratios are not computable from one another.

Impact on RD, OR or RR if you...

	....interchange	
	Exposed/Not Exposed	Case/Control
RD	sign change	sign change
OR	reciprocal	reciprocal
RR	reciprocal	not straightforward

Dependence on the measure [RD, OR or RR]

Now we can see that an interpretation of results can depend on the measure to be used in assessing an intervention.

Rate differences can offer very different interpretations from rate ratios.

Odds ratios can offer a different message from rate ratios.

A rate ratio based on failure can provide a different message from a rate ratio based on success.

We can also see the importance of ‘looking at the data’ and not just looking at the summary measures.

What now?

Based on the simple analysis, one might have concluded that the overall success rate could be improved if treatment 2 were always used.

In fact, we can see that treatment 1 has a higher success rate for both surgery types.

Here, a ‘stratified’ analysis may have enabled a ‘salvage job’.

There are a number of choices. They should be made IN ADVANCE at the design stage.

Using software:

Using Stata

```
. cs suc tr

                 | tr                     |
                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+------------
           Cases |       195         505  |        700
        Noncases |       905         595  |       1500
-----------------+------------------------+------------
           Total |      1100        1100  |       2200
                 |                        |
            Risk |  .1772727    .4590909  |   .3181818
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
 Risk difference |         -.2818182      |     -.31892   -.2447163
      Risk ratio |          .3861386      |    .3348359    .4453018
                 +-------------------------------------------------
                               chi2(1) =   201.35  Pr>chi2 = 0.0000
```

```
. gen fail=1-suc

. cs fail tr

                 | tr                     |
                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+------------
           Cases |       905         595  |       1500
        Noncases |       195         505  |        700
-----------------+------------------------+------------
           Total |      1100        1100  |       2200
                 |                        |
            Risk |  .8227273    .5409091  |   .6818182
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
 Risk difference |         .2818182       |    .2447163      .31892
      Risk ratio |         1.521008       |    1.431053    1.616618
                 +-------------------------------------------------
                                  chi2(1) =   201.35  Pr>chi2 = 0.0000
```

Rate Difference and Rate Ratio

RD(S) is estimated by - 0.2818

and RD(F) is estimated by 0.2818

RD(F) = - RD(S)

RR(S) is estimated by 0.3861

and RR(F) is estimated by 1.521

You cannot compute RR(F) from RR(S) alone and vice versa

p-value ($p<0.001$ and not $p=0.0000$) refers to the approximate χ^2 test. Fisher's exact test can be reported using the exact option

```
. cc suc tr
                                                         Proportion
                 |   Exposed   Unexposed  |      Total      Exposed
-----------------+------------------------+------------------------
           Cases |       195         505  |        700       0.2786
        Controls |       905         595  |       1500       0.6033
-----------------+------------------------+------------------------
           Total |      1100        1100  |       2200       0.5000
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         .2538701       |    .2077658    .3099697 (exact)
                 +-------------------------------------------------
                               chi2(1) =   201.35  Pr>chi2 = 0.0000

. cc fail tr
                                                         Proportion
                 |   Exposed   Unexposed  |      Total      Exposed
-----------------+------------------------+------------------------
           Cases |       905         595  |       1500       0.6033
        Controls |       195         505  |        700       0.2786
-----------------+------------------------+------------------------
           Total |      1100        1100  |       2200       0.5000
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         3.939022       |    3.226121    4.813098 (exact)
                 +-------------------------------------------------
                               chi2(1) =   201.35  Pr>chi2 = 0.0000
```

Odds Ratios

OR(S) is estimated by 0.2539

and OR(F) is estimated by 3.939 = 1/0.2539

OR(F)=1/OR(S)

OR(F) can be computed from OR(S) alone and vice versa

computing an estimate of RR(F)

1.521 = 3.939*0.3861 = 0.3861/0.2539

notice that RR(F) is not 1/ RR(S)

```
. cs suc tr,by(surg)

            surg |       RR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               1 |          2      .8342841    4.79453     4.545455
               2 |        1.9      1.759944   2.051202     45.45455
-----------------+-------------------------------------------------
           Crude |   .3861386      .3348359   .4453018
    M-H combined |   1.909091       1.71543   2.124615
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    0.026  Pr>chi2 = 0.8724

. cs fail tr,by(surg)

            surg |       RR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               1 |   .9473684        .90163   .9954271     86.36364
               2 |         .1      .0424614   .2355078     45.45455
-----------------+-------------------------------------------------
           Crude |   1.521008      1.431053   1.616618
    M-H combined |   .6551724       .580039   .7400379
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =  231.866  Pr>chi2 = 0.0000
```

Rate Ratio analyses are very different

the stratum specific estimates are as computed [by hand] earlier

the assessment of modification depends whether one considers success or failure

the adjusted estimates have differing roles:

1.909 might be included in a final summary since the stratum specific estimates are so close

but 0.6552 would have no place in such a summary given that the stratum specific estimates are so different

```
. cc suc tr,by(surg)

            surg |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               1 |   2.111111      .8436801   6.804342     4.090909 (exact)
               2 |         19      7.753503   60.26036     2.272727 (exact)
-----------------+-------------------------------------------------
           Crude |   .2538701      .2077658   .3099697              (exact)
    M-H combined |   8.142857      4.342777   15.26814
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    11.57  Pr>chi2 = 0.0007

                   Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =     67.85
                                                Pr>chi2 =    0.0000
. cc fail tr,by(surg)

            surg |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               1 |   .4736842      .1469826   1.184903     8.636364 (exact)
               2 |   .0526316      .0165967   .1289397     43.18182 (exact)
-----------------+-------------------------------------------------
           Crude |   3.939022      3.226121   4.813098              (exact)
    M-H combined |    .122807      .0654959   .2302674
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    11.57  Pr>chi2 = 0.0007

                   Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =     67.85
                                                Pr>chi2 =    0.0000
```

Odds Ratio analyses yield the same results

simply recall that $OR(F) = 1/OR(S)$

then all estimates can be directly compared noting the reciprocal relationship

All Six Counterfactual Plots

We can view and consider all six of the plots of the counterfactuals.

RD(S) and RD(F) contain the same information as do OR(S) and OR(F).

RR(S) and RR(F) are not the same.

Setting up counterplot.ado in Stata

First, identify your personal adopath directory

In Stata, type :

personal

Then, add counterplot.ado to your personal adopath directory. You may need to create the directory first.

After adding the .ado file, check that all is well by typing :

personal dir

Stata [on a Mac] responds with :

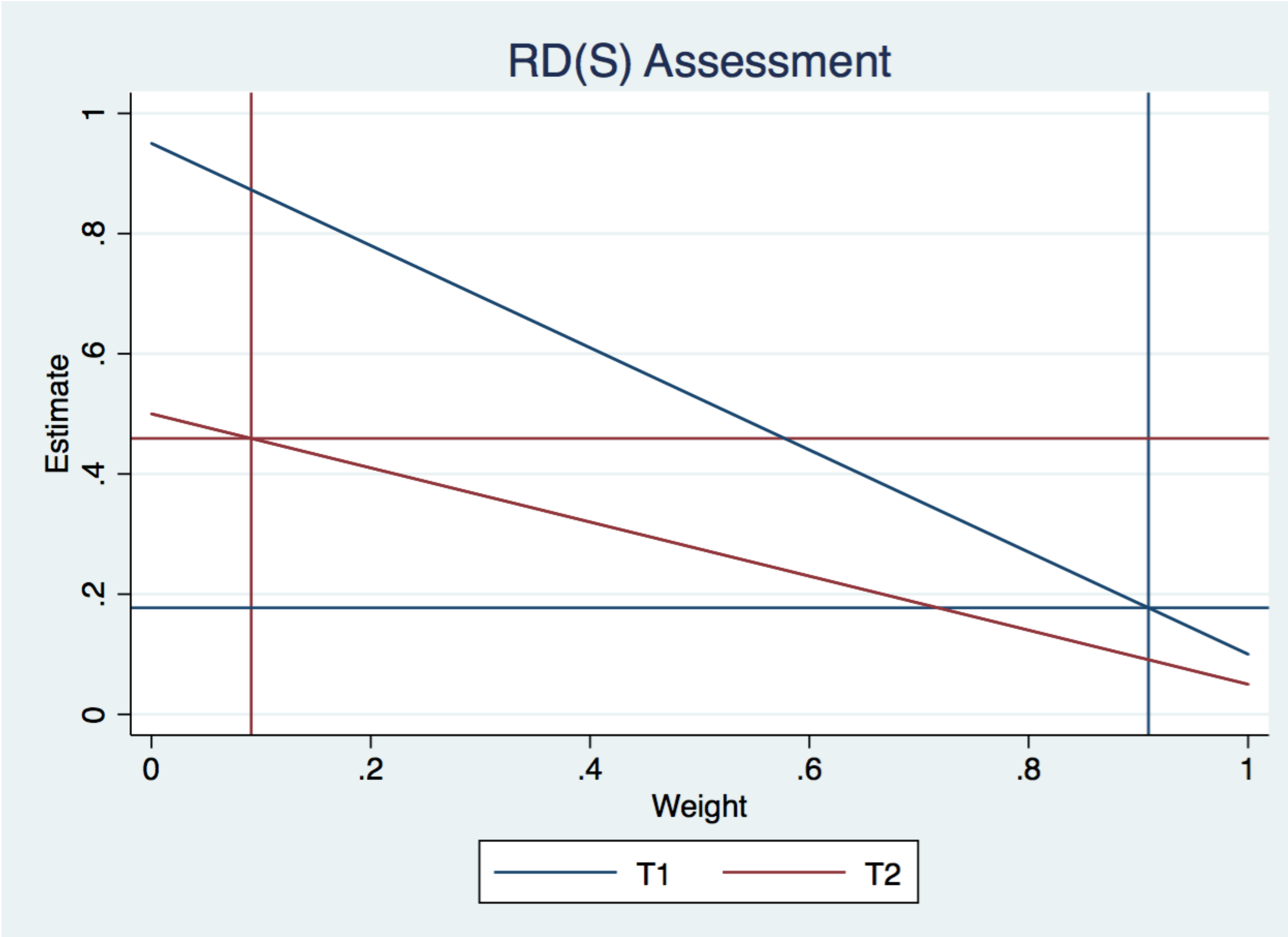
your personal ado-directory is

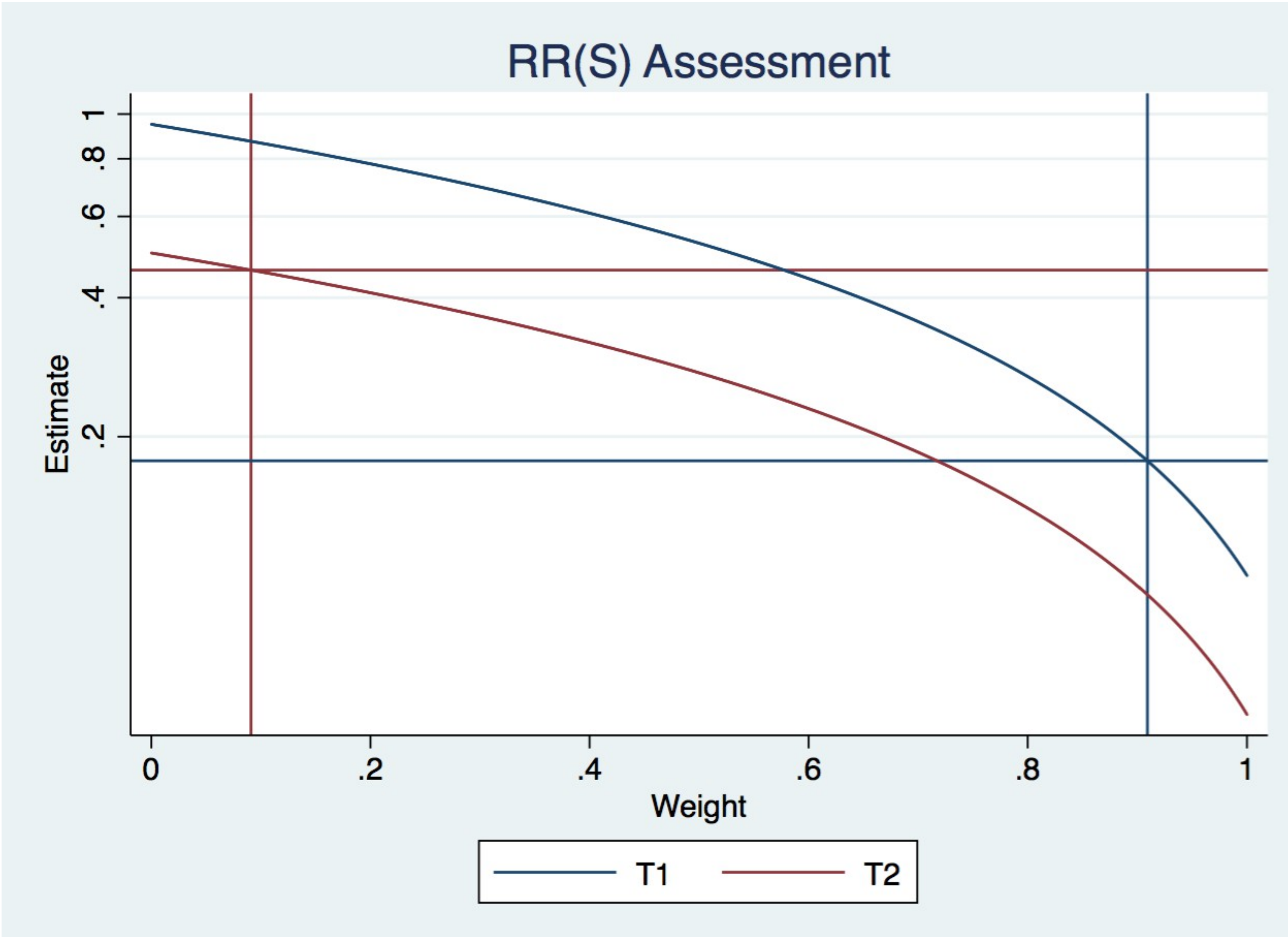
~/Library/Application Support/Stata/ado/personal/

counterplot.ado

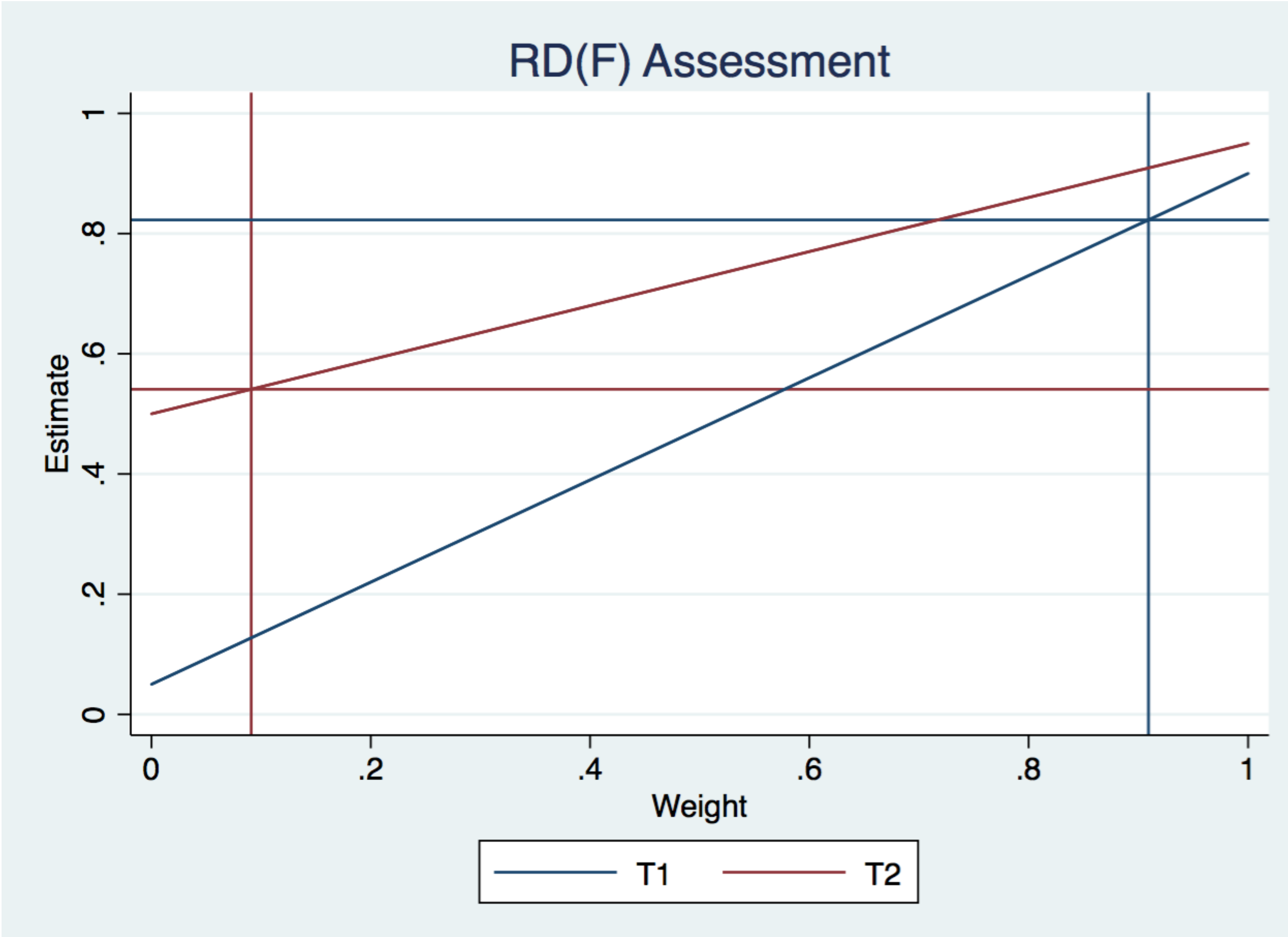
Using counterplot.ado

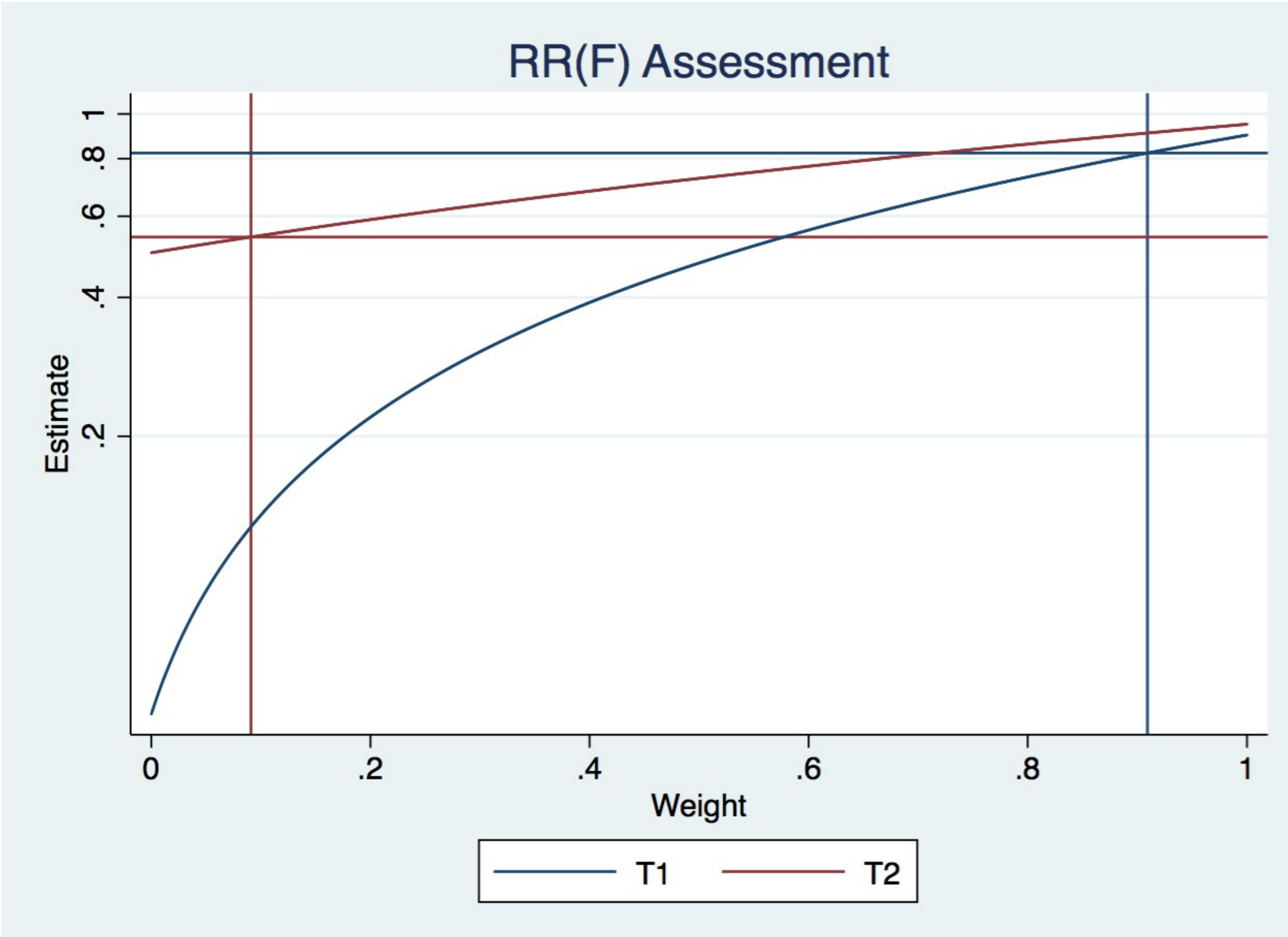
```
use kalbfleisch_p180.dta

counterplot fail tr surg

counterplot fail tr surg,measure(rr)

counterplot fail tr surg,measure(or)

counterplot fail tr surg,measure(rr) outcome(success)
```

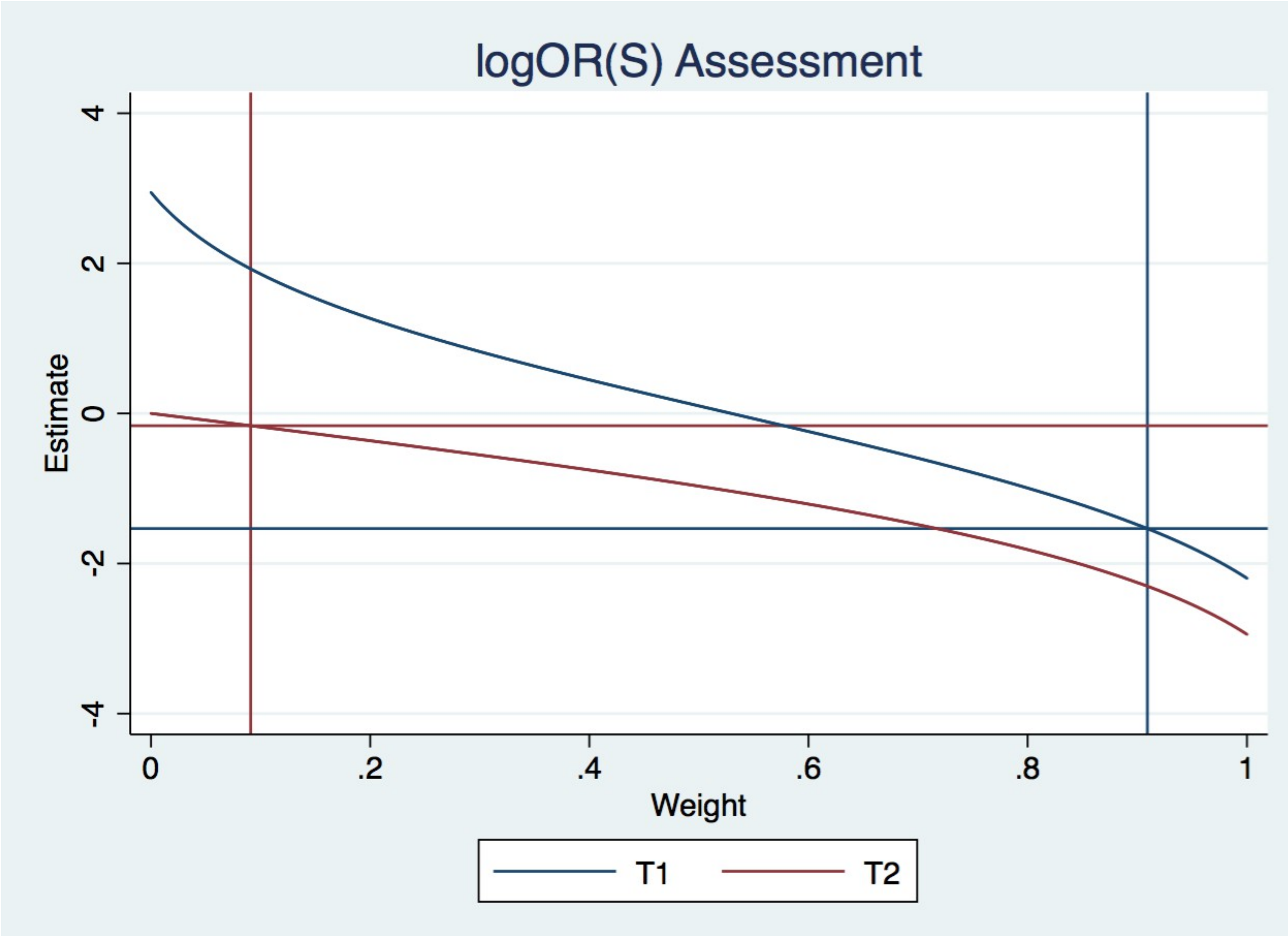

RD(S) Assessment
Estimate
0
.2
.4
.6
.8
1
Weight
T1
T2



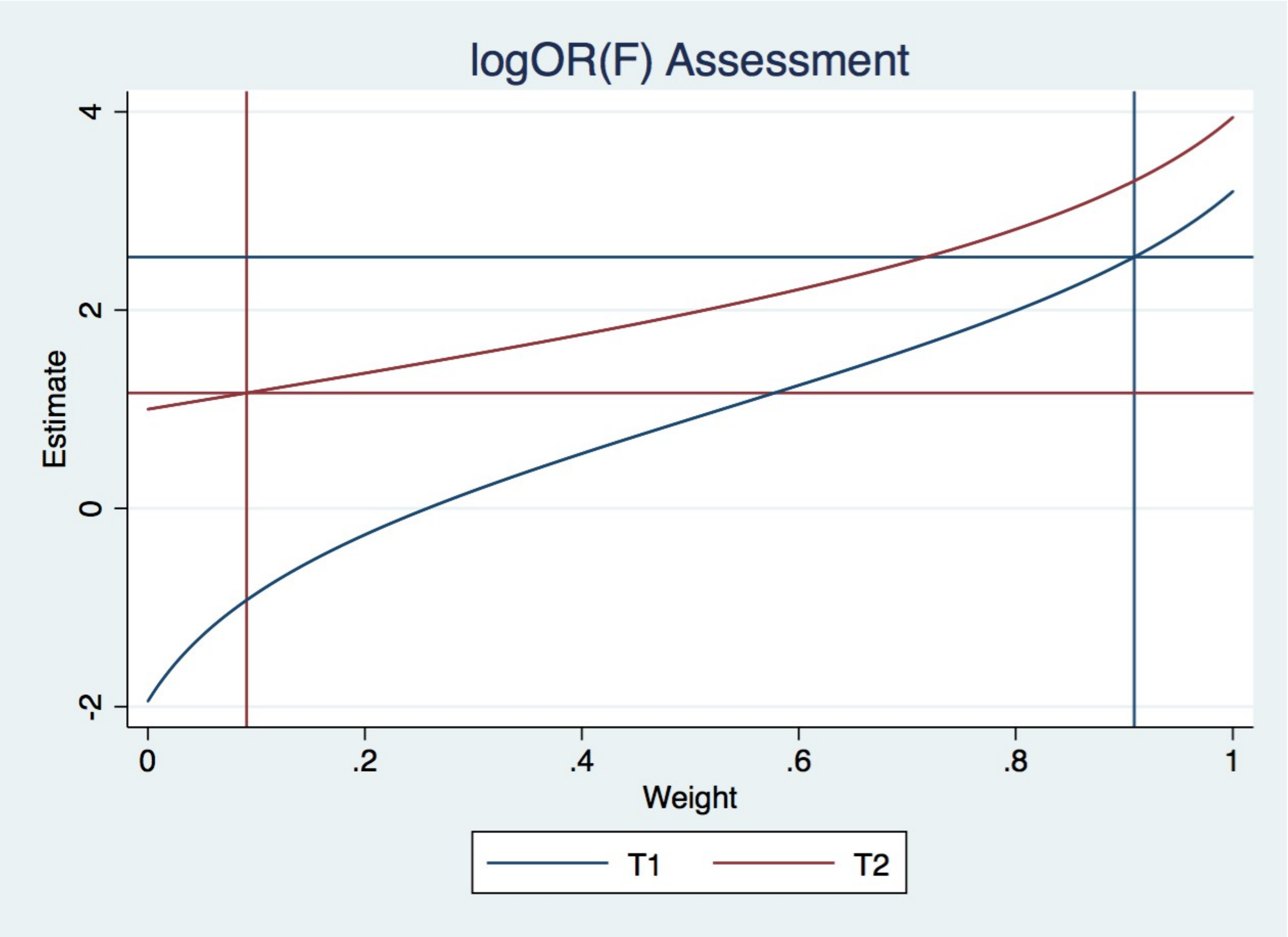
RR(S) Assessment
Estimate
1
.8
.6
.4
.2
0
.2
.4
.6
.8
1
Weight
T1
T2


RD(F) Assessment
Estimate
0
.2
.4
.6
.8
1
Weight
0
.2
.4
.6
.8
1
T1
T2


RR(F) Assessment
Estimate
.2
.4
.6
.8
1
0
.2
.4
.6
.8
1
Weight
T1
T2



logOR(S) Assessment
Estimate
4
2
0
-2
-4
0
.2
.4
.6
.8
1
Weight
T1
T2


logOR(F) Assessment
Estimate
4
2
0
-2
0
.2
.4
.6
.8
1
Weight
T1
T2

Some formatting tips

When pasting Stata output into a word processor, it is best to a fixed spacing font like Courier.

Courier 9 point Bold fits on most pages and gives the clearest look.

Stata graphs seem clearest inside a word processor if they are imported in .png format

Use .gph format if you want to return to a graph within Stata