Biost 536: Categorical Data Analysis in Epidemiology
Emerson, Autumn 2013

Homework #1
September 26, 2013

Written problems due at 5 pm, Thursday, October 3, 2013. Homeworks must be submitted electronically according to the instructions that will be distributed via email.

This homework explores the role of screening studies in promoting the accuracy of the process of identifying and quantifying risk factors for disease.

The goal of the drug approval process should be

1. To have a low probability of approving drugs that do not work,
2. To have a high probability of approving drugs that do work, and
3. To have a high probability that an approved drug does work.

Now suppose we decide to perform a experiment or series of experiments, and to approve the drug whenever the estimated treatment effect (perhaps standardized to some Z score) exceeds a pre-defined threshold. When stated in statistical jargon, these goals become

1. To have a low type I error α when a null hypothesis of no treatment effect is true,
2. To have a high statistical power $Pwr = 1-\beta$ (so β is the type II error) when some alternative hypothesis is true, and
3. To have a high positive predictive value PPV = (number of approved effective drugs) / (number of approved drugs).

We can examine the interrelationships of these statistical design criteria in the context of a RCT where we let θ denote our treatment effect, and we presume that an ineffective drug has $\theta = 0$, and an effective drug has some $\theta > 0$.

In the "frequentist" inference most often used in RCT, we typically choose some value for the "level of significance" (or type I error) α. This will be the probability of approving the drug when $\theta = 0$.

Most often, we base our decisions on some estimate of the treatment effect that is known to be approximately normally distributed

$$\hat{\theta} \sim N\left(\theta, \frac{V}{n}\right).$$

In experimental design, we sometimes choose a sample size n and then compute the power of the study to detect a particular alternative hypothesis. When our null hypothesis corresponds to $\theta = 0$, the power of a particular design depends upon the type I error α, the variability of the data V, the true value of the treatment effect θ, and the sample size n according to the following formula:

$$Pwr = 1 - \Pr\left(Z \le z_{1-\alpha} - \theta\sqrt{\frac{n}{V}}\right), \qquad \text{(Eq. 1)}$$

where Z is a random variable having the standard normal distribution, and the constant $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution such that $\Pr(Z \le z_{1-\alpha}) = 1 - \alpha$.

In other settings, we choose a desired power *Pwr* = 1 - β, and then compute a sample size according to the value of β using the following formula (which again presumes a null hypothesis of θ = 0):

$$n = \frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2 V}{\theta^2}, \qquad \text{(Eq. 2)}$$

where we again use the quantiles of the standard normal distribution. The following table provides values of $z_{1-\alpha}$ for selected values of α:

α	0.005	0.01	0.025	0.05	0.10	0.20
$z_{1-\alpha}$	2.575829	2.326348	1.959964	1.644854	1.281552	0.841621

More generally, we can obtain an arbitrary quantile using statistical software. The commands to obtain the $z_{1-\alpha}$ quantile when α = 0.075 in three commonly used programs are:

- (Stata) `di invnorm(1 - 0.075)`
- (R) `qnorm(1 - 0.075)`
- (Excel) `norminv(1 - 0.075, 0 , 1)`

Similarly, we can obtain Pr($Z \leq c$) for arbitrary choices of *c* using statistical software. The commands to obtain Pr($Z \leq c$) when *c* = 1.75 in three commonly used programs are:

- (Stata) `di norm(1.75)`
- (R) `pnorm(1.75)`
- (Excel) `normdist(1.75, 0 , 1, TRUE)`

Bayes Rule can be used to compute the *PPV* from α and β, providing we know the prior probability π that a treatment would work (this prior probability might be thought of as the proportion of effective treatments among all treatments that we would consider testing—sort of a prevalence of good treatments):

$$PPV = \frac{(1-\beta)\times\pi}{(1-\beta)\times\pi + \alpha\times(1-\pi)} \qquad \text{(Eq. 3)}$$

In this homework, we consider a couple examples of two different strategies of testing for experimental treatments:

1. Strategy 1: Test each treatment in one large "pivotal" RCT.
2. Strategy 2: Test each treatment in one small "pilot" RCT that screens for promising treatments. Any treatment that passes this screening phase, is then tested more rigorously in one larger "confirmatory" RCT.

To compare "apples with apples":

- We pretend that we have 500,000 patients with disease X to use when evaluating ideas that we have formulated for treating disease X.
- We further pretend that 10% of our ideas correspond to drugs that truly work (so π = 0.10), and all those truly effective drugs provide the same degree of benefit θ = 1 to patients with disease X. The other 90% of our ideas correspond to drugs that provide no benefit to the patients (so θ = 0).
- In every RCT, the true variability of the patient data corresponds to *V* = 63.70335.

PLEASE SEE LAST PAGES FOR DETAILED CALCULATIONS USED

Problems using Strategy 1: Only Pivotal RCT

1. (A: Pivotal) Suppose we choose a type I error of $\alpha = 0.025$ and a power of 97.5% (so $\beta = 0.025$) under the alternative hypothesis that the true treatment effect is $\theta = 1$.
   a. What sample size *n* will be used in each RCT? **979**

$$n = \frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2 V}{\theta^2} = \frac{\left(1.959964 + 1.959964\right)^2 \times 63.70335}{1^2} = 978.855$$

   b. How many of our ideas will we be able to test? **511**

      500,000 / 979 = 510.7

   c. How many of those tested ideas will be truly beneficial drugs? **51**

      511 x 0.10 = 51.1

   d. How many of the tested beneficial drugs will have significant results? **50**

      51 x 0.975 = 49.7

   e. How many of those tested ideas will be truly ineffective drugs? **460**

      511 – 51 = 460

   f. How many of the tested ineffective drugs will have significant results? **12**

      460 x 0.025 = 11.5

   g. How many of the tested drugs will have significant results? **62**

      50 + 12 = 62

   h. What proportion of the drugs with significant results will be truly beneficial? **0.8065**

      50 / 62 = 0.8065 or

$$PPV = \frac{(1-\beta)\times\pi}{(1-\beta)\times\pi + \alpha\times(1-\pi)} = \frac{(1-0.025)\times 0.10}{(1-0.025)\times 0.10 + 0.025\times(1-0.10)} = 0.8125$$

2. (B: Pivotal) Suppose we choose a type I error of $\alpha = 0.025$ and a power of 80.0% (so $\beta = 0.20$) under the alternative hypothesis that the true treatment effect is $\theta = 1$.
   a. What sample size *n* will be used in each RCT? **500**
   b. How many of our ideas will we be able to test? **1000**
   c. How many of those tested ideas will be truly beneficial drugs? **100**
   d. How many of the tested beneficial drugs will have significant results? **80**
   e. How many of those tested ideas will be truly ineffective drugs? **900**
   f. How many of the tested ineffective drugs will have significant results? **23**
   g. How many of the tested drugs will have significant results? **103**

h. What proportion of the drugs with significant results will be truly beneficial? **0.78**

3. (C: Pivotal) Suppose we choose a type I error of $\alpha = 0.05$ and a power of 80.0% (so $\beta = 0.20$) under the alternative hypothesis that the true treatment effect is $\theta = 1$.
   a. What sample size *n* will be used in each RCT? **394**
   b. How many of our ideas will we be able to test? **1269**
   c. How many of those tested ideas will be truly beneficial drugs? **127**
   d. How many of the tested beneficial drugs will have significant results? **102**
   e. How many of those tested ideas will be truly ineffective drugs? **1142**
   f. How many of the tested ineffective drugs will have significant results? **57**
   g. How many of the tested drugs will have significant results? **159**
   h. What proportion of the drugs with significant results will be truly beneficial? **0.64**

Problems using Strategy 2: Screening pilot RCT, followed by Confirmatory RCT

4. (D: Screening pilot study) Suppose we choose a type I error of $\alpha = 0.025$ and a sample size of $n = 100$ for each pilot RCT.
   a. Under the alternative hypothesis $\theta = 1$, what is the power? **0.29**
   b. If we use 350,000 patients in pilot RCT, how many ideas will we test? **3500**
   c. How many of those tested ideas will be truly beneficial drugs? **350**
   d. How many of the tested beneficial drugs will have significant results? **103**
   e. How many of those tested ideas will be truly ineffective drugs? **3150**
   f. How many of the tested ineffective drugs will have significant results? **79**
   g. How many of the tested drugs will have significant results? **182**
   h. What proportion of the drugs with significant results will be truly beneficial? **0.56**

5. (D: Confirmatory trials) Suppose we choose a type I error of $\alpha = 0.025$ and use all remaining patients in the confirmatory trials of each drug that had significant results in problem 4.
   a. How many confirmatory RCT will be performed? **824**
   b. What sample size *n* will be used in each RCT? **182**
   c. Under the alternative hypothesis $\theta = 1$, what is the power? **0.73**
   d. How many confirmatory RCTs will be for truly beneficial drugs? **82**
   e. How many of the tested beneficial drugs will have significant results? **60**
   f. How many confirmatory RCTs will be for truly ineffective drugs? **742**
   g. How many of the tested ineffective drugs will have significant results? **19**
   h. How many of the tested drugs will have significant results? **79**
   i. What proportion of the drugs with significant results will be truly beneficial? **0.76**

6. (E: Screening pilot study) Suppose we choose a type I error of $\alpha = 0.10$ and a power of 85.0% (so $\beta = 0.15$) under the alternative hypothesis that the true treatment effect is $\theta = 1$.
   a. What sample size *n* will be used in each RCT? **342**
   b. If we use 350,000 patients in pilot RCT, how many ideas will we test? **1023**
   c. How many of those tested ideas will be truly beneficial drugs? **102**
   d. How many of the tested beneficial drugs will have significant results? **87**
   e. How many of those tested ideas will be truly ineffective drugs? **921**
   f. How many of the tested ineffective drugs will have significant results? **92**
   g. How many of the tested drugs will have significant results? **179**
   h. What proportion of the drugs with significant results will be truly beneficial? **0.49**

7. (E: Confirmatory trials) Suppose we choose a type I error of $\alpha = 0.025$ and use all remaining patients in the confirmatory trials of each drug that had significant results in problem 6.
   a. How many confirmatory RCT will be performed? **838**
   b. What sample size *n* will be used in each RCT? **179**
   c. Under the alternative hypothesis $\theta = 1$, what is the power? **0.72**
   d. How many confirmatory RCTs will be for truly beneficial drugs? **84**
   e. How many of the tested beneficial drugs will have significant results? **60**
   f. How many confirmatory RCTs will be for truly ineffective drugs? **754**
   g. How many of the tested ineffective drugs will have significant results? **19**
   h. How many of the tested drugs will have significant results? **79**
   i. What proportion of the drugs with significant results will be truly beneficial? **0.76**

Comparisons

8. Of the 5 different strategies considered (problems 1, 2, 3, 4 and 5, or 6 and 7) which do you think best and why?

   The three goals of the drug approval process are: 1) increasing the probability of approving drugs which work (sensitivity); 2) decreasing the probability of approving drugs which do not work (specificity); and 3) increasing the probability of approved drugs working (positive predictive value). In addition, we always want to ensure drug safety and limit our use of resources (increase efficiency). Given these goals, I believe that starting with a pilot study followed by a confirmatory trial is the best approach. Pilot studies are small and help decide which drugs are promising and should be investigated further, which is a good use of resources. Pilot studies also provide initial safety information before exposing larger numbers of subjects to potentially harmful effects. For the confirmatory trials, we want to maximize sensitivity, specificity, and the positive predictive value. Unfortunately, there is always a tradeoff between these measures, so the goal is to balance and maximize each as much as possible.

As such, I believe that the trials described in problems 4 and 5 is the best approach. This approach will result in conducting 350 RCTs in the pilot phase and 824 RCTs in the confirmatory phase. The sensitivity and specificity in the confirmatory trials are (specificity=1-alpha=.975) (sensitivity=1-beta=.73) 0.73 and 0.975 respectively, and the positive predictive value is 0.76. These results are similar to the pilot and confirmatory phases of questions 6 and 7; however the number of trials required for the confirmatory phase is larger (less efficiency).

9. The above exercises considered "drug discovery" with randomized clinical trials. What additional issues have to be considered when we are using observational data to explore and try to confirm risk factors for particular diseases?

The two largest issues with observational studies are confounding and the placebo effect. Randomization allows us to assume that both observed and unobserved confounders are equally distributed between the groups (in most cases). In observational studies, randomization is not possible and although there are methods that we can utilize to control or adjust for potential confounders statistically, we can only use these methods for observable confounders which results in residual confounding from the unobserved confounders. The placebo effect is an issue for accurately determining a treatment effect. When the subject is not blinded, or knows what their treatment assignment is, they can think that they feel better on the active therapy or don't feel better on the placebo regardless of whether the therapy is effective or not. It is important to use blinding to avoid this type of bias, and blinding is typically not possible in observational studies.

Biost 536: Categorical Data Analysis in Epidemiology

Emerson, Autumn 2013

Homework #1

September 26, 2013

Written problems due at 5 pm, Thursday, October 3, 2013. Homeworks must be submitted electronically according to the instructions that will be distributed via email.

This homework explores the role of screening studies in promoting the accuracy of the process of identifying and quantifying risk factors for disease.

The goal of the drug approval process should be

1. To have a low probability of approving drugs that do not work,
2. To have a high probability of approving drugs that do work, and
3. To have a high probability that an approved drug does work.

Now suppose we decide to perform a experiment or series of experiments, and to approve the drug whenever the estimated treatment effect (perhaps standardized to some Z score) exceeds a pre-defined threshold. When stated in statistical jargon, these goals become

1. To have a low type I error α when a null hypothesis of no treatment effect is true,
2. To have a high statistical power $Pwr = 1-\beta$ (so β is the type II error) when some alternative hypothesis is true, and
3. To have a high positive predictive value PPV = (number of approved effective drugs) / (number of approved drugs).

We can examine the interrelationships of these statistical design criteria in the context of a RCT where we let θ denote our treatment effect, and we presume that an ineffective drug has $\theta = 0$, and an effective drug has some $\theta > 0$.

In the "frequentist" inference most often used in RCT, we typically choose some value for the "level of significance" (or type I error) α. This will be the probability of approving the drug when $\theta = 0$.

Most often, we base our decisions on some estimate of the treatment effect that is known to be approximately normally distributed

$$\hat{\theta} \sim N\left(\theta, \frac{V}{n}\right).$$

In experimental design, we sometimes choose a sample size n and then compute the power of the study to detect a particular alternative hypothesis. When our null hypothesis corresponds to $\theta = 0$, the power of a particular design depends upon the type I error α, the variability of the data V, the true value of the treatment effect θ, and the sample size n according to the following formula:

$$Pwr = 1 - \Pr\left(Z \le z_{1-\alpha} - \theta\sqrt{\frac{n}{V}}\right), \qquad \text{(Eq. 1)}$$

where Z is a random variable having the standard normal distribution, and the constant $z_{1-\alpha}$ is the 1-α quantile of the standard normal distribution such that $\Pr(Z \le z_{1-\alpha}) = 1 - \alpha$.

In other settings, we choose a desired power $Pwr = 1 - \beta$, and then compute a sample size according to the value of β using the following formula (which again presumes a null hypothesis of $\theta = 0$):

$$n = \frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2 V}{\theta^2}, \qquad \text{(Eq. 2)}$$

where we again use the quantiles of the standard normal distribution. The following table provides values of $z_{1-\alpha}$ for selected values of α:

α	0.005	0.01	0.025	0.05	0.10	0.20
$z_{1-\alpha}$	2.575829	2.326348	1.959964	1.644854	1.281552	0.841621

More generally, we can obtain an arbitrary quantile using statistical software. The commands to obtain the $z_{1-\alpha}$ quantile when $\alpha = 0.075$ in three commonly used programs are:

- (Stata) `di invnorm(1 - 0.075)`
- (R) `qnorm(1 - 0.075)`
- (Excel) `norminv(1 - 0.075, 0 , 1)`

Similarly, we can obtain $\Pr(Z \le c)$ for arbitrary choices of c using statistical software. The commands to obtain $\Pr(Z \le c)$ when $c = 1.75$ in three commonly used programs are:

- (Stata) `di norm(1.75)`
- (R) `pnorm(1.75)`
- (Excel) `normdist(1.75, 0 , 1, TRUE)`

Bayes Rule can be used to compute the PPV from α and β, providing we know the prior probability π that a treatment would work (this prior probability might be thought of as the proportion of effective treatments among all treatments that we would consider testing—sort of a prevalence of good treatments):

$$PPV = \frac{(1-\beta)\times\pi}{(1-\beta)\times\pi + \alpha\times(1-\pi)} \qquad \text{(Eq. 3)}$$

In this homework, we consider a couple examples of two different strategies of testing for experimental treatments:

1. Strategy 1: Test each treatment in one large "pivotal" RCT.
2. Strategy 2: Test each treatment in one small "pilot" RCT that screens for promising treatments. Any treatment that passes this screening phase, is then tested more rigorously in one larger "confirmatory" RCT.

To compare "apples with apples":

- We pretend that we have 500,000 patients with disease X to use when evaluating ideas that we have formulated for treating disease X.
- We further pretend that 10% of our ideas correspond to drugs that truly work (so $\pi = 0.10$), and all those truly effective drugs provide the same degree of benefit $\theta = 1$ to patients with disease X. The other 90% of our ideas correspond to drugs that provide no benefit to the patients (so $\theta = 0$).
- In every RCT, the true variability of the patient data corresponds to $V = 63.70335$.

Power = 1 − β
1 − specificity

Problems using Strategy 1: Only Pivotal RCT

1. (A: Pivotal) Suppose we choose a type I error of $\alpha = 0.025$ and a power of 97.5% (so $\beta = 0.025$) under the alternative hypothesis that the true treatment effect is $\theta = 1$.
   a. What sample size *n* will be used in each RCT? **979**

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 V}{\theta^2} = \frac{(1.959964 + 1.959964)^2 \times 63.70335}{1^2} = 978.855$$

   b. How many of our ideas will we be able to test? **511**

   500,000 / 979 = 510.7

   c. How many of those tested ideas will be truly beneficial drugs? **51**

   511 x 0.10 = 51.1

   d. How many of the tested beneficial drugs will have significant results? **50**

   51 x 0.975 = 49.7

   e. How many of those tested ideas will be truly ineffective drugs? **460**

   511 – 51 = 460

   f. How many of the tested ineffective drugs will have significant results? **12**

   460 x 0.025 = 11.5

   g. How many of the tested drugs will have significant results? **62**

   50 + 12 = 62

   h. What proportion of the drugs with significant results will be truly beneficial? **0.8065**

   50 / 62 = 0.8065 or

$$PPV = \frac{(1-\beta)\times\pi}{(1-\beta)\times\pi + \alpha\times(1-\pi)} = \frac{(1-0.025)\times 0.10}{(1-0.025)\times 0.10 + 0.025\times(1-0.10)} = 0.8125$$

V = 63.70335
π = .10

0.0975
0.0975 + .025 = .0975
0.12 = .812

2. (B: Pivotal) Suppose we choose a type I error of $\alpha = 0.025$ and a power of 80.0% (so $\beta = 0.20$) under the alternative hypothesis that the true treatment effect is $\theta = 1$.
   a. What sample size *n* will be used in each RCT? 500
   b. How many of our ideas will we be able to test? 500,000 / 500 = 1000 (# of RCTs)
   c. How many of those tested ideas will be truly beneficial drugs? (1000)(.10) = 100
   d. How many of the tested beneficial drugs will have significant results? (100)(.80) = 80
   e. How many of those tested ideas will be truly ineffective drugs? 900
   f. How many of the tested ineffective drugs will have significant results? 23
   g. How many of the tested drugs will have significant results? 103
   h. What proportion of the drugs with significant results will be truly beneficial? 0.78

1000 − 100 = 900 OR (1000)(.90) = 900
(900)(.025) = 22.5
80 + 23 =
80/103 = .777

$$PPV = \frac{(1-.20)*0.10}{(1-.20)*0.10 + 0.025\times(1-.10)} = \frac{.08}{.08 + .0225} = \frac{.08}{.1025} = .78$$

3. (C: Pivotal) Suppose we choose a type I error of $\alpha = 0.05$ and a power of 80.0% (so $\beta = 0.20$) under the alternative hypothesis that the true treatment effect is $\theta = 1$.
   a. What sample size n will be used in each RCT? 394
   b. How many of our ideas will we be able to test? $500{,}000/394 =$ 1269
   c. How many of those tested ideas will be truly beneficial drugs? 127
   d. How many of the tested beneficial drugs will have significant results? 102
   e. How many of those tested ideas will be truly ineffective drugs? 1142
   f. How many of the tested ineffective drugs will have significant results? 57
   g. How many of the tested drugs will have significant results? 159
   h. What proportion of the drugs with significant results will be truly beneficial? 0.64

$n = \frac{(1.644854 + .8416212)^2 (63.70335)}{12} = (6.18557726)(63.70335) = 393.85$

$(1269)(.10) = 126.9$ $(127)(.80) = 101.5$ $(1269)(.90) = 1142.1$ $(1142)(0.05) = 57.1$

$PPV = \frac{(1-.20)(.10)}{(.08) + (.05)(.90)} = \frac{.08}{.125} = 0.64$

Problems using Strategy 2: Screening pilot RCT, followed by Confirmatory RCT

4. (D: Screening pilot study) Suppose we choose a type I error of $\alpha = 0.025$ and a sample size of $n = 100$ for each pilot RCT.
   a. Under the alternative hypothesis $\theta = 1$, what is the power? 0.293
   b. If we use 350,000 patients in pilot RCT, how many ideas will we test? $= 350{,}000/100 = 3500$ (# RCTs) 3500
   c. How many of those tested ideas will be truly beneficial drugs? $= (3500)(.10) = 350$ 350
   d. How many of the tested beneficial drugs will have significant results? $= (350)(0.293) = 102.55$ 103
   e. How many of those tested ideas will be truly ineffective drugs? $= (3500)(.90) = 3150$ 3150
   f. How many of the tested ineffective drugs will have significant results? $= (3150)(.025) = 78.75$ 79
   g. How many of the tested drugs will have significant results? $= 103 + 79 - 182$ 182
   h. What proportion of the drugs with significant results will be truly beneficial? $= 103/182 = .566$ 0.56

$PPV = \frac{(1-.707)(.10)}{(1-.707)(.10) + (.025)(.90)} = .293$

$1-\alpha = 0.975$ $z_{1-\alpha} = 1.959964$ $1-(1.959964 - (1)\sqrt{100}/63.70335) = 1 - 1.252107 = -.252107$

$1 - .7056917 = 0.2943083$

5. (D: Confirmatory trials) Suppose we choose a type I error of $\alpha = 0.025$ and use all remaining patients in the confirmatory trials of each drug that had significant results in problem 4.
   a. How many confirmatory RCT will be performed? $500{,}000 - 350{,}000 = 150{,}000/182 = 824.2$ 824
   b. What sample size n will be used in each RCT? $150{,}000/824 = 182$ 182
   c. Under the alternative hypothesis $\theta = 1$, what is the power? .73
   d. How many confirmatory RCTs will be for truly beneficial drugs? $= (824)(.10) = 82.4$ ↙ power 82
   e. How many of the tested beneficial drugs will have significant results? $= (82)(.73) = 59.86$ 60
   f. How many confirmatory RCTs will be for truly ineffective drugs? $= (824)(.90) = 741.6$ 742
   g. How many of the tested ineffective drugs will have significant results? $= (742)(.025) = 18.55$ 19
   h. How many of the tested drugs will have significant results? $= 60 + 19 = 79$ 79
   i. What proportion of the drugs with significant results will be truly beneficial? $60/79 = .759$ .759

$\pi = 0.79$ $\alpha = .025$

$1-(1.959964 - (1)\sqrt{182}/63.7335) = 1 - 1.690268004 = 1 - (.2690994) = 0.7309$

6. (E: ~~Screening pilot~~ study) Suppose we choose a type I error of $\alpha = 0.10$ and a power of 85.0% (so $\beta = 0.15$) under the alternative hypothesis that the true treatment effect is $\theta = 1$.
   a. What sample size n will be used in each RCT? 342
   b. If we use 350,000 patients in pilot RCT, how many ideas will we test? 1023
   c. How many of those tested ideas will be truly beneficial drugs? 102
   d. How many of the tested beneficial drugs will have significant results? 87
   e. How many of those tested ideas will be truly ineffective drugs? 921
   f. How many of the tested ineffective drugs will have significant results? 92
   g. How many of the tested drugs will have significant results? 179
   h. What proportion of the drugs with significant results will be truly beneficial? 0.486

7. (E: Confirmatory trials) Suppose we choose a type I error of $\alpha = 0.025$ and use all remaining patients in the confirmatory trials of each drug that had significant results in problem 6.
   a. How many confirmatory RCT will be performed? 838
   b. What sample size n will be used in each RCT? 179
   c. Under the alternative hypothesis $\theta = 1$, what is the power? .716
   d. How many confirmatory RCTs will be for truly beneficial drugs? 84
   e. How many of the tested beneficial drugs will have significant results? 60
   f. How many confirmatory RCTs will be for truly ineffective drugs? 754
   g. How many of the tested ineffective drugs will have significant results? 19
   h. How many of the tested drugs will have significant results? 79
   i. What proportion of the drugs with significant results will be truly beneficial? .76

Comparisons

8. Of the 5 different strategies considered (problems 1, 2, 3, 4 and 5, or 6 and 7) which do you think best and why?

9. The above exercises considered "drug discovery" with randomized clinical trials. What additional issues have to be considered when we are using observational data to explore and try to confirm risk factors for particular diseases?