Sample Size and Power I: Binary Outcomes

James Ware, PhD
Harvard School of Public Health
Boston, MA

Sample Size and Power

Principles:

Sample size calculations are an essential part of study design

Consider sample size requirements early

A well-designed trial is large enough to detect clinically important differences between groups with high probability

To perform sample size calculations, we need well defined study endpoints, hypotheses, and statistical tests.

Specify the Null Hypothesis

Study hypotheses should be based on a clearly defined endpoint and period of study.

In most RCTs, known as superiority trials, the study hypothesis is stated as a null hypothesis of no difference in the distribution of the primary endpoint between study groups.

In the CORONARY Trial, the short-term null hypothesis was

H_0: Patients receiving on-pump and off-pump coronary artery bypass surgery will have identical event rates at 30 days post-randomization

Specify the Alternative Hypothesis

We have an alternative hypothesis in mind, for example,

H_A: The frequency of events at 30 days will differ in the two treatment groups.

In superiority trials, we test the null hypothesis against a two-sided alternative.

We have a directional alternative hypothesis in mind, for example, that fewer events will occur within 30 days in the off-pump group.

Outcomes of Hypothesis Testing

When we test the null hypothesis, there are two possible states of nature and two decisions:

	Truth About Risk Difference	
Test Result	H_0 True	H_a True
Reject H_0	Type I Error	No Error
Do Not Rej H_0	No Error	Type II Error

Power

We will perform a test that has a small probability of a Type 1 error, usually 0.05.

The power of the study is the probability that we will reject the null hypothesis when the alternative hypothesis is actually true.

We would like this probability to be large, typically at least 0.8.

Express the Hypotheses in Terms of Probabilities

The 30-day outcome is a binary event, occurrence or non-occurrence of death or complications within 30 days of surgery.

p_T = Probability that an off-pump patient will have an event

p_C = Probability that an on-pump patient will have an event

The study hypotheses are:

H_0: $p_T = p_C$ (T and C are equally effective)

H_A: $p_T \neq p_C$ (T and C are not equally effective)

We specify the direction of the alternative for the sample size calculation.

The Test Statistic

To test the null hypothesis, we calculate a test statistic, T, and a critical value, C, and reject the null hypothesis if $|T| > C$, that is, if $T > C$ or $T < -C$.

To calculate power or sample size, we will focus on significance in one direction, $T < -C$, implying that $p_T < p_C$.

For the CORONARY Trial, define T as the difference between the observed proportions divided by the standard error of the difference.

The Test Statistic

The observed difference in proportions is $D = \hat{p}_T - \hat{p}_C$

Assuming equal sample sizes in the two groups,

$$\text{Var}(D) = p_T*(1-p_T)/n + p_C*(1-p_C)/n$$

Under the null hypothesis, $p_T = p_C$. Define the test statistic as D divided by its standard deviation

$$T = \frac{\hat{p}_T - \hat{p}_C}{\sqrt{2\bar{p}*(1-\bar{p})/n}} = \frac{D}{SD(D)}$$

where $\bar{p}$ is the average event rate

Choosing C

Choose C so that $P(T < -C \mid H_0) = \alpha/2$. Usually, $\alpha = 0.05$ (two-sided) so $\alpha/2 = 0.025$.

T is approximately N(0,1) if H_0 is true. Hence, if

$$\alpha/2 = 0.025, C = 1.96.$$

Power is $P(T < -C \mid H_A) = 1 - P(\text{Type 2 error}) = 1 - \beta$.

The investigator can control the power by choosing the sample size

Null and Alternative Hypothesis

In the CORONARY Trial, one possible scenario for the 30-day endpoint was

$p_C = 0.08$, and, under the alternative hypothesis

$p_T = (0.85)*0.08 = 0.068$

a 15% reduction in the event rate in the off-pump group. Under H_A, the expected value of D would be

$$\Delta = 0.068 - 0.08 = -0.012$$

If H_0 is true, var(D) = 2*0.08*0.92/n

Logic of Sample Size Calculations

Again consider the risk difference, $D = \hat{p}_T - \hat{p}_C$

D is approximately normally distributed with

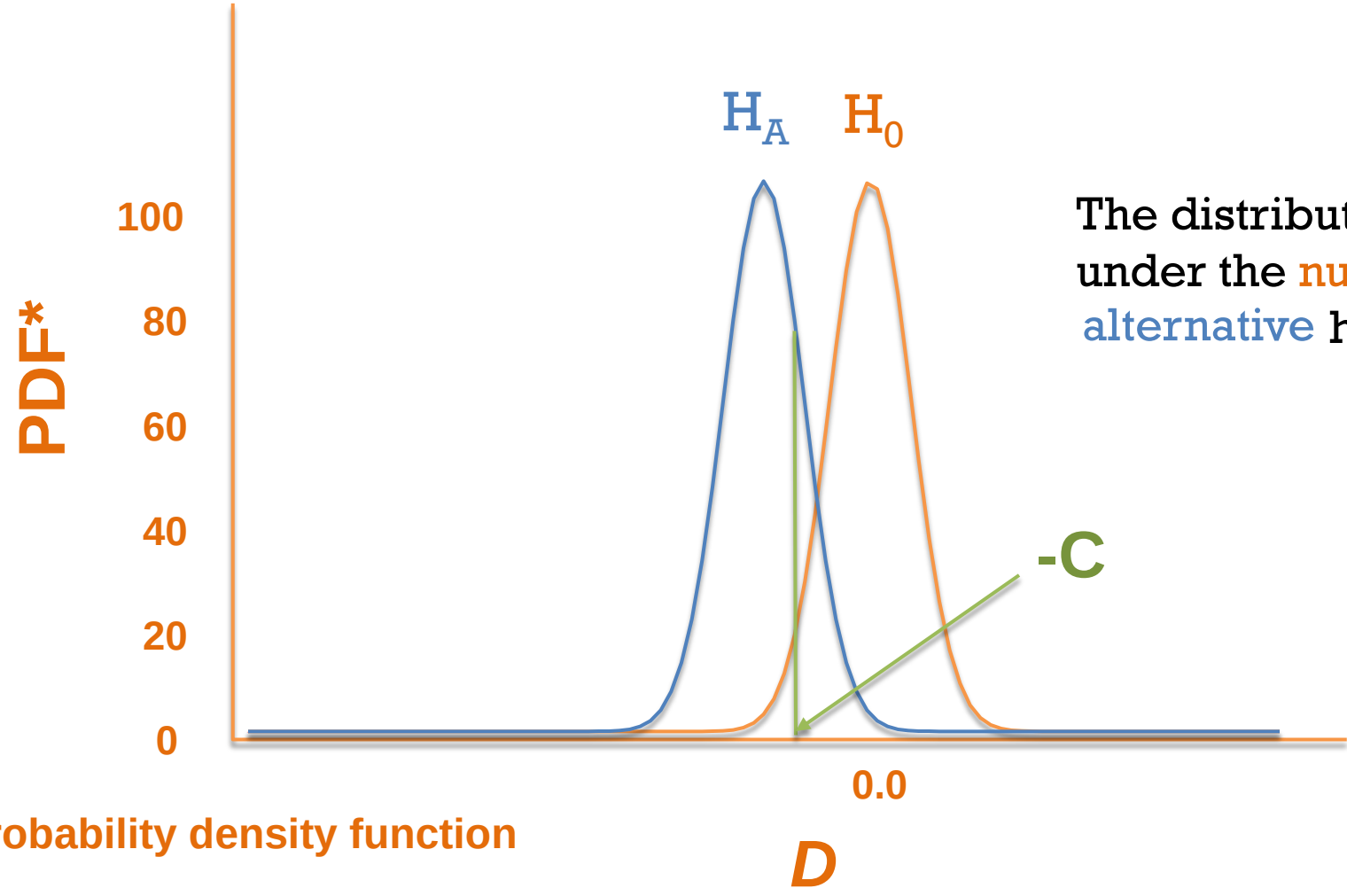
mean = 0 if H_0 is true and

mean = -.012 if H_A is true

$$\text{Var}(D) = p_T*(1-p_T)/n_T + p_C*(1-p_C)/n_C$$

The mean is independent of n but the variance decreases as n increases

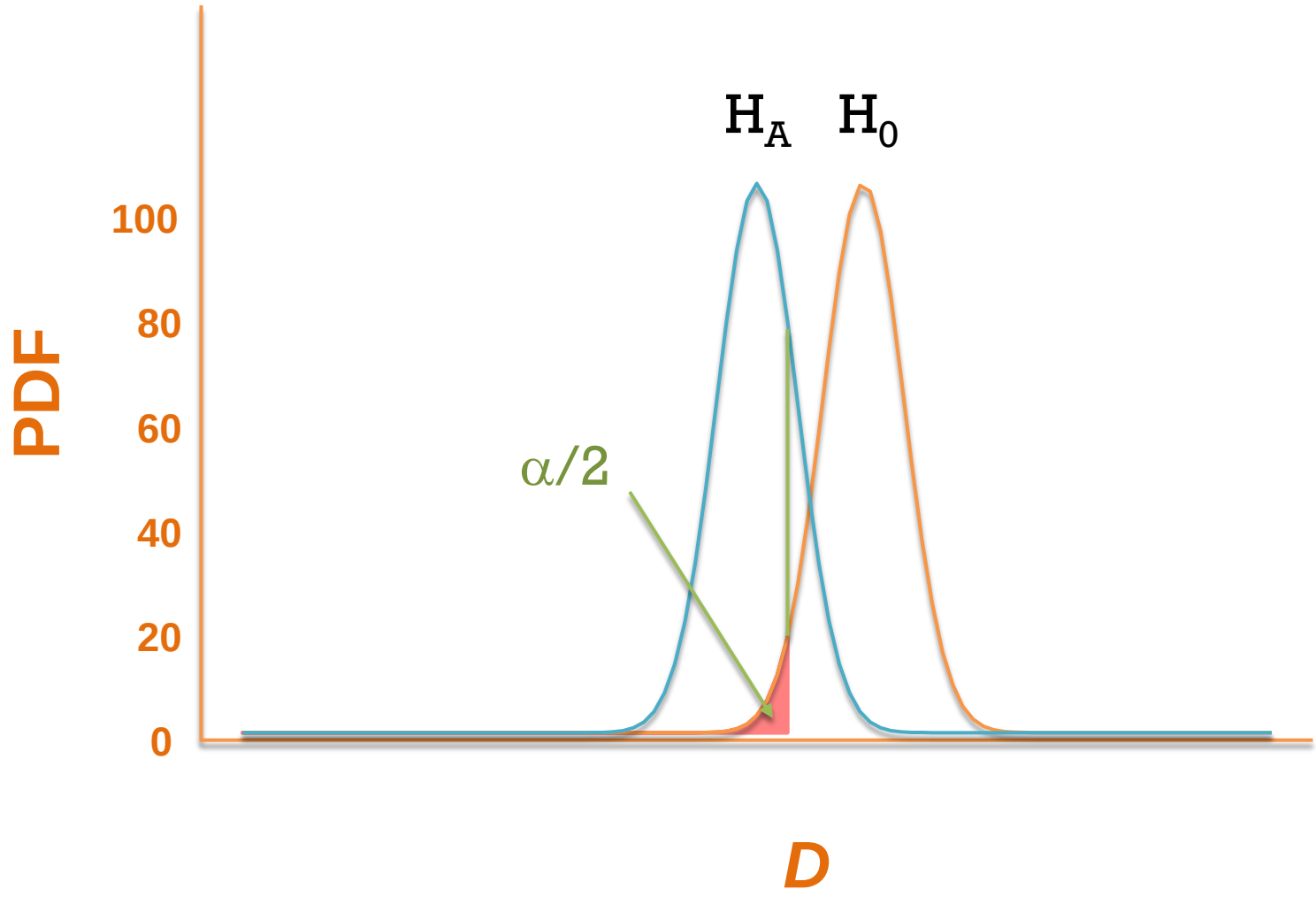
The Logic of Hypothesis Testing



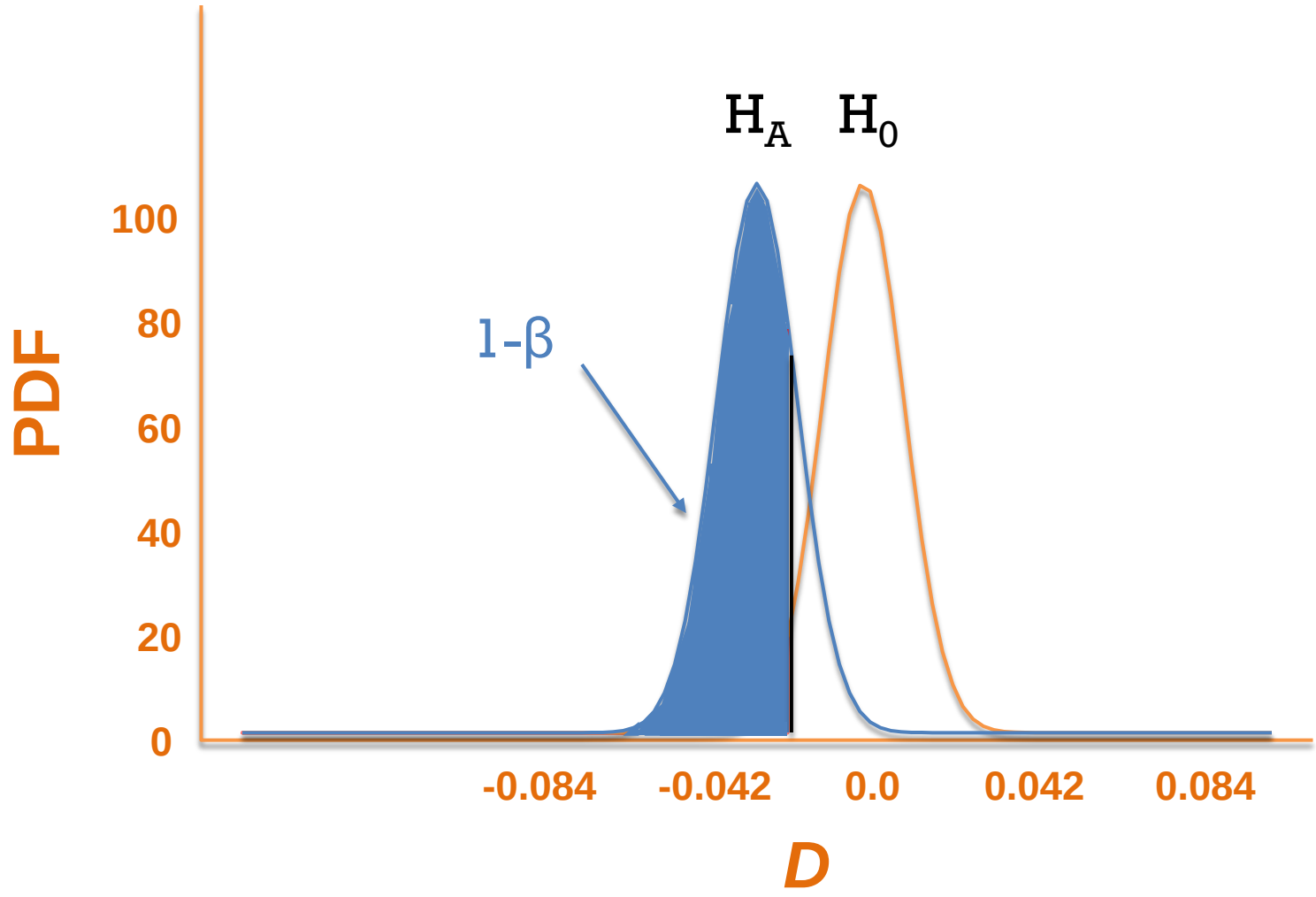
The distribution of *D* under the null and alternative hypotheses

*pdf: probability density function

Type 1 Error

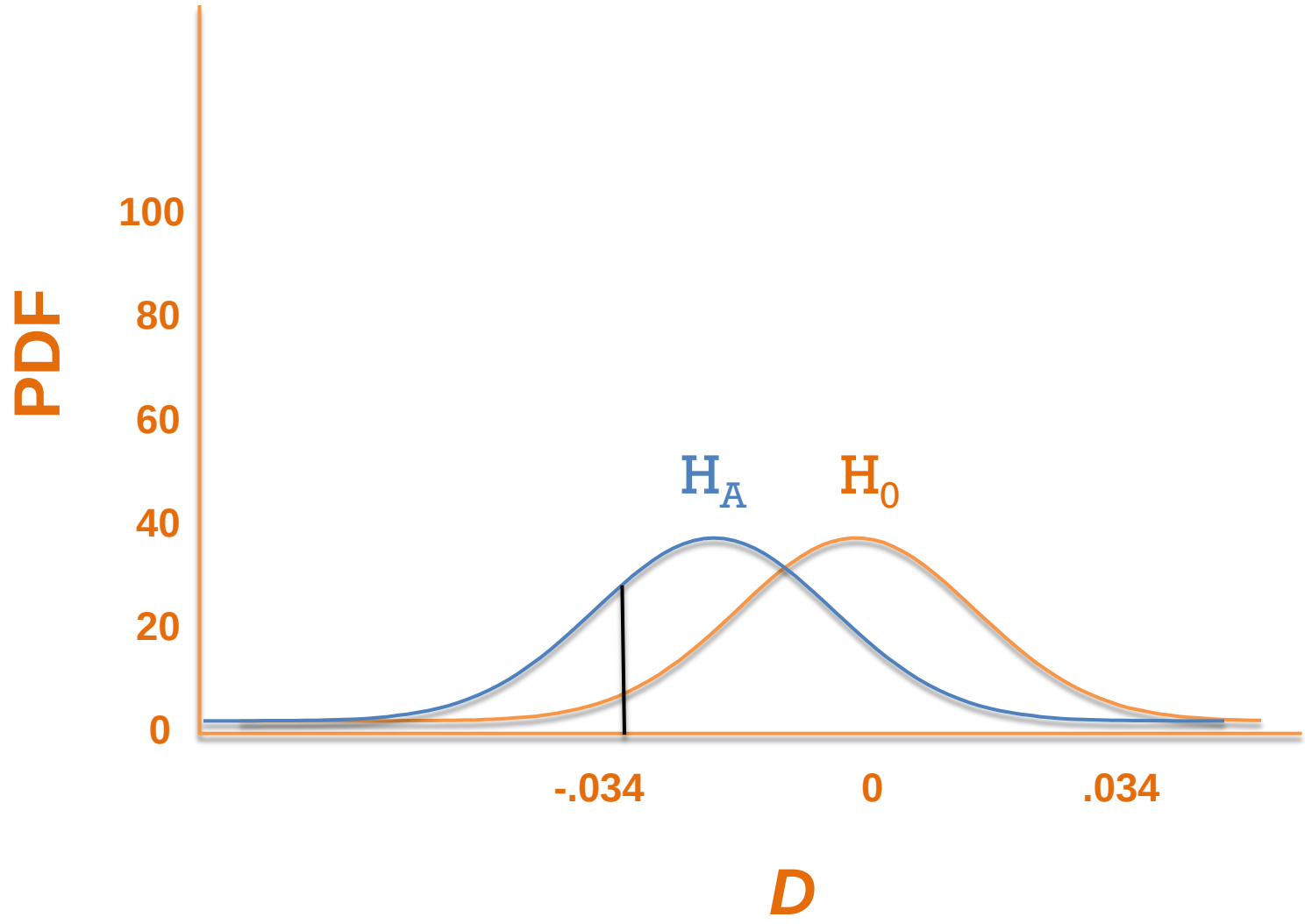


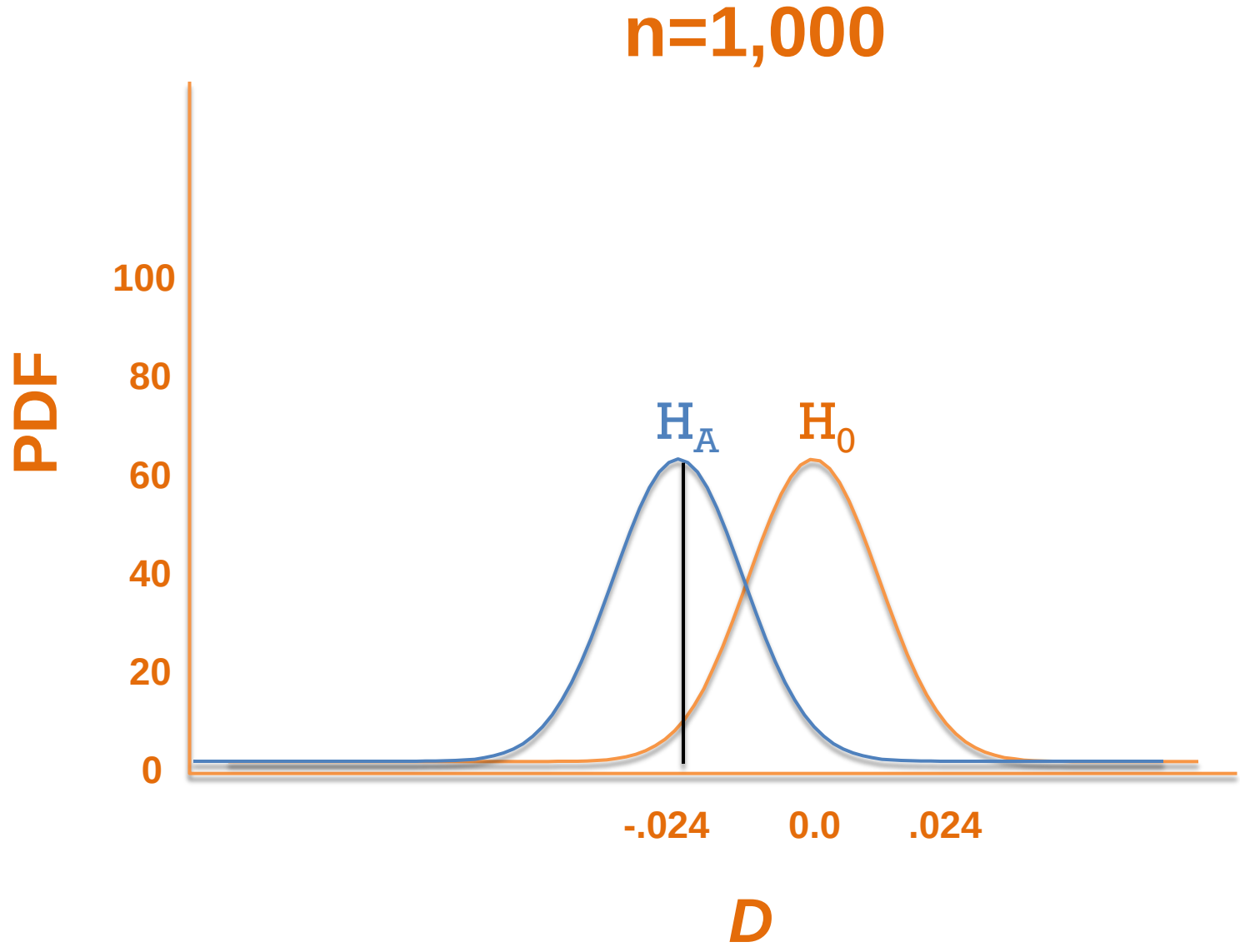
Power

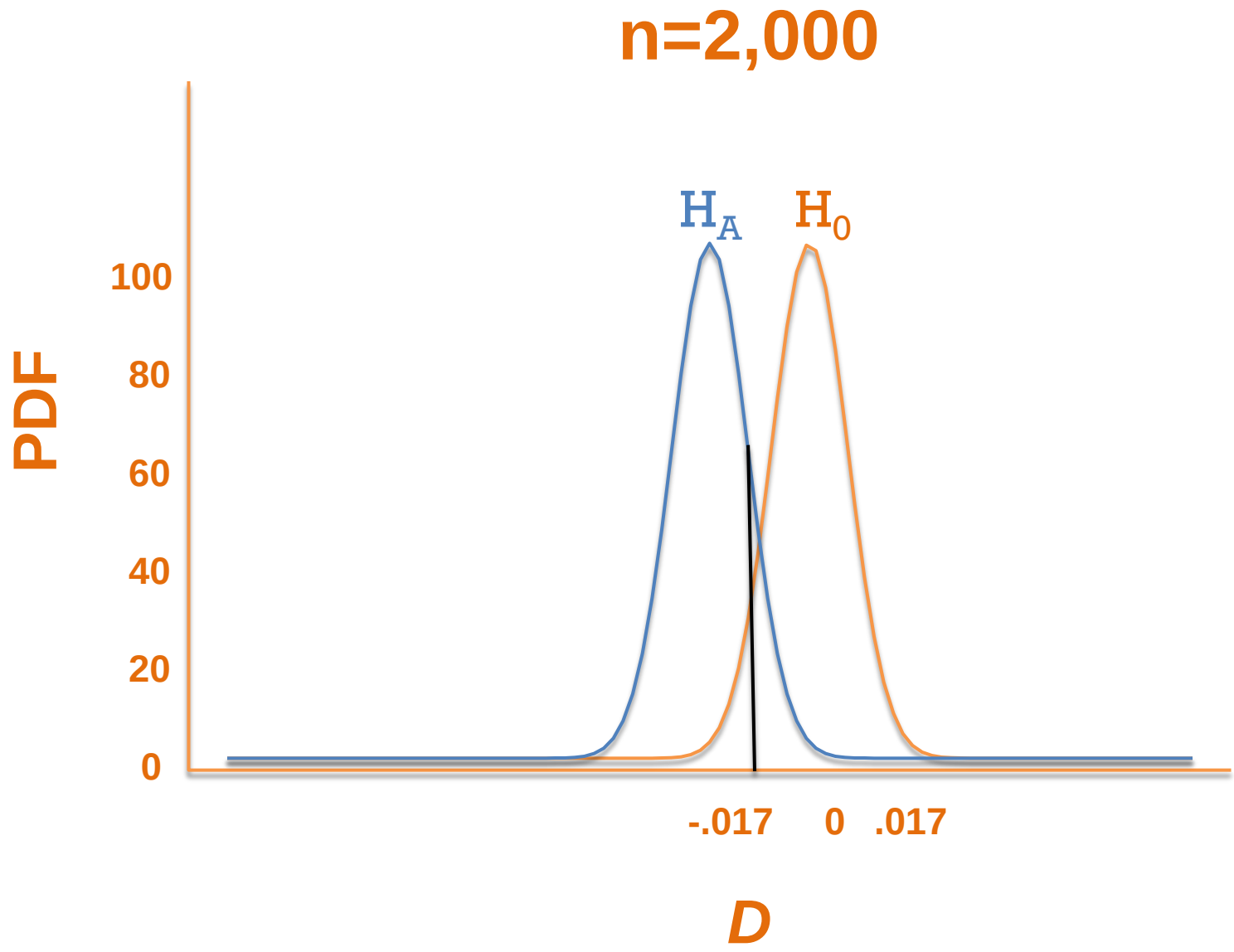


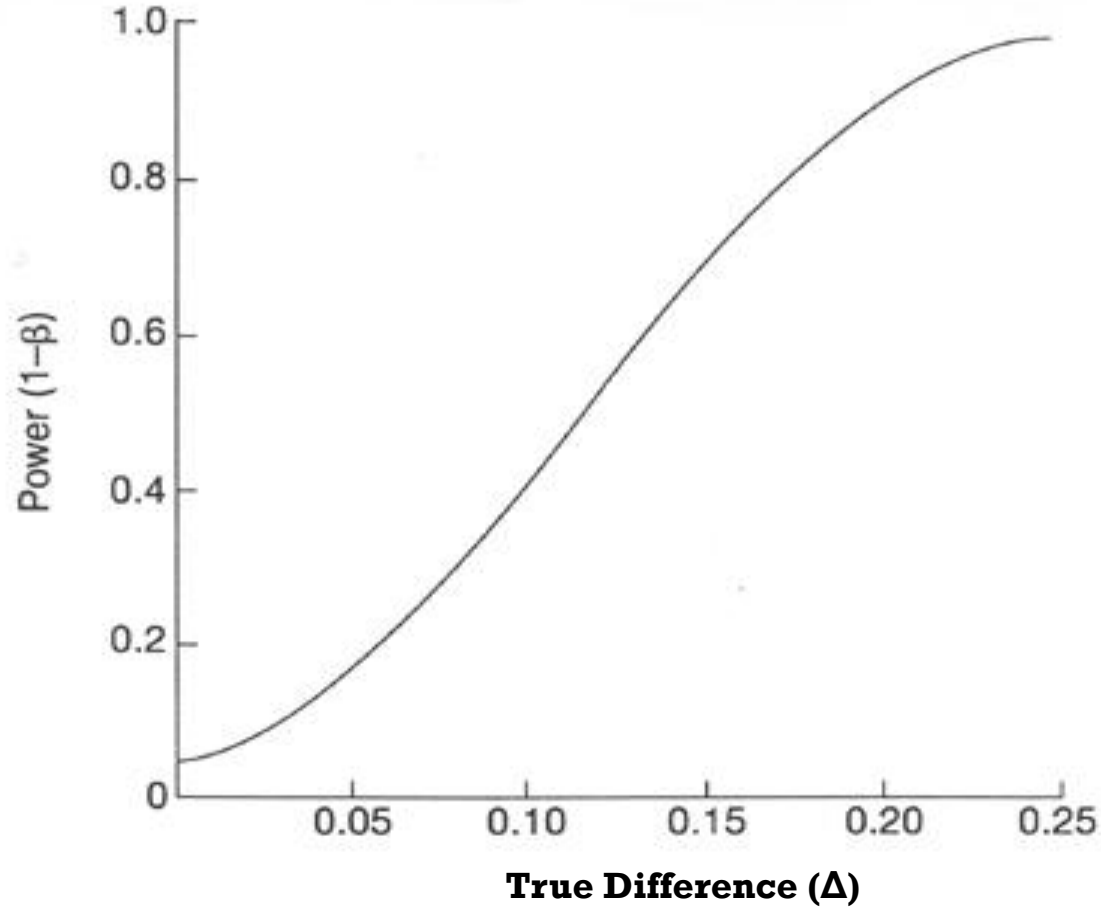
Var(*D*) Depends on Sample Size

n = 500	Var = 0.00029	SD = 0.017
n = 1,000	Var = 0.00015	SD = 0.012
n = 2,000	Var = 0.000074	SD = 0.0086


n= 500
PDF
100
80
60
40
20
0
H_A
H_0
-.034
0
.034
D


n=1,000
PDF
100
80
60
40
20
0
H_A
H_0
-.024
0.0
.024
D


n=2,000
H_A
H_0
PDF
100
80
60
40
20
0
-.017
0
.017
D



For a fixed sample size, the power of the study will increase with the size of the true difference

Sample Size Formula

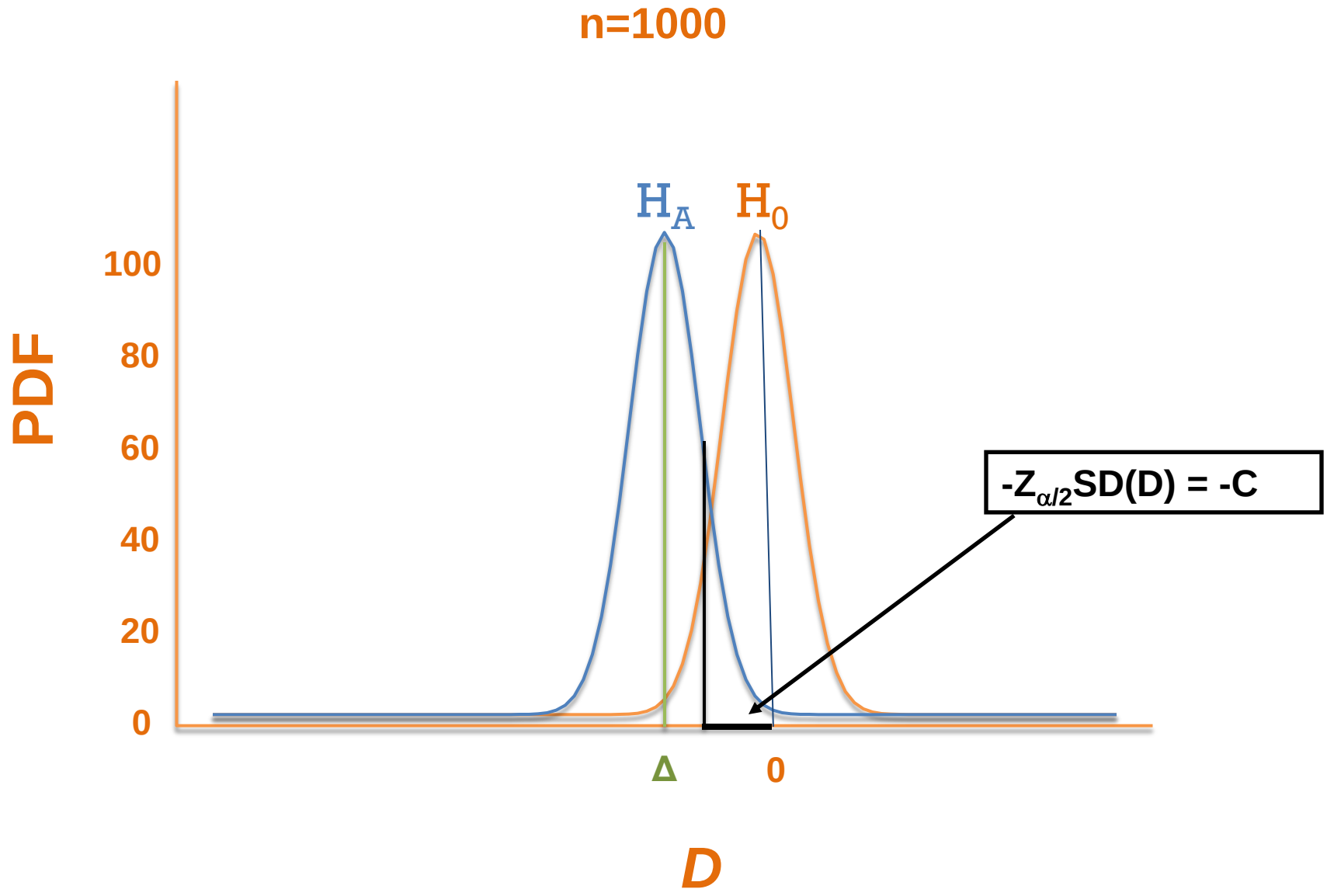
To achieve the desired Type 1 and Type 2 error, we need to satisfy two conditions

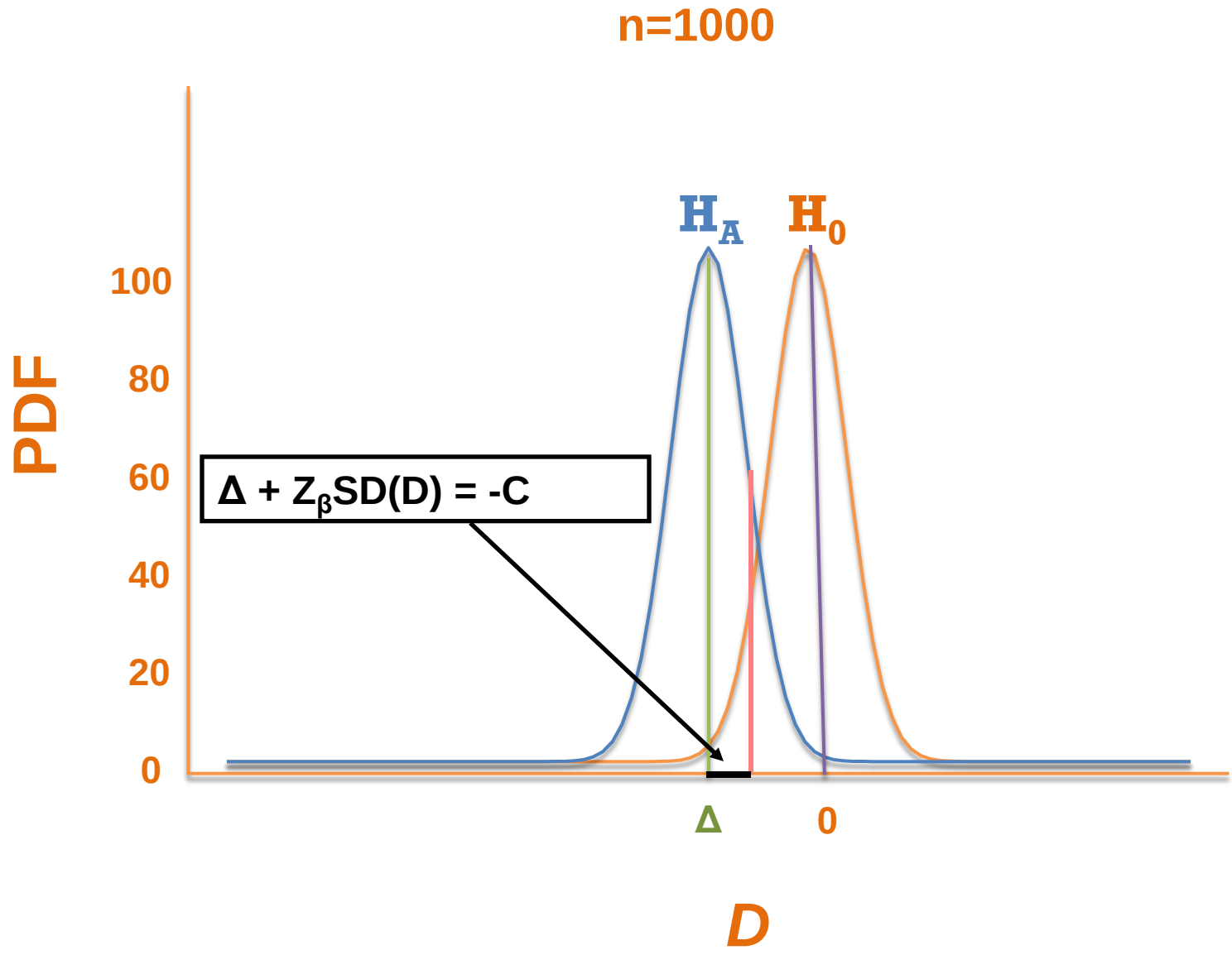
$$-Z_{\alpha/2}*SD(D) = -C \text{ and } \Delta + Z_{\beta}*SD(D) = -C$$

Recall that we estimate the variance of D by

$$2\bar{p} * (1 - \bar{p})/n$$

To determine n, set $-Z_{\alpha/2}*SD(D) = \Delta + Z_{\beta}*SD(D)$ and solve for n


n=1000
PDF
100
80
60
40
20
0
H_A
H_0
$-Z_{\alpha/2}SD(D) = -C$
Δ
0
D


n=1000
PDF
100
80
60
40
20
0
H_A
H_0
$\Delta + Z_\beta SD(D) = -C$
Δ
0
D

The Sample Size Formula

$$n = \frac{2\bar{p} * (1 - \bar{p})\left(Z_{\alpha/2} + Z_{\beta}\right)^2}{\Delta^2}$$

$Z_{\alpha/2}$ and Z_{β} are the critical values of the normal distribution, $\bar{p}$ is the average of the event rates under the alternative hypothesis, and Δ is the true difference under H_A. For the CORONARY Trial, with

$$\alpha = 0.05, \beta = 0.20, n = 1{,}903 \text{ or } 2n = 3{,}806.$$

(Note: If $p_c = 0.08$, and $p_t = 0.068$ if H_A is true, the correct value for the sample size is n = 7,462.)

The CORONARY investigators considered a range of scenarios and settled on a total sample size of 4,700 patients.