Sample Size, Power and Sampling Methods

Mary Ann McBurnie, PhD
Senior Investigator, Kaiser Permanente Center for Health Research
Steering Committee Chair, Community Health Applied Research Network (CHARN)

Jonathan N. Tobin, PhD
Professor, Albert Einstein College of Medicine of Yeshiva University
President/CEO CLINICAL DIRECTORS NETWORK, INC. (CDN)

CHARN Community Health Applied Research Network

KAISER PERMANENTE®

The Center for Health Research

EnCoRE Presentation, Feb 17, 2015

Acknowledgment

Material in this presentation was developed as part of the curriculum for the international **Methods in Epidemiologic, Clinical and Operations Research (MECOR)** program sponsored by the **American Thoracic Society (ATS)**

- Designed for physicians and health care professionals
- Intended to strengthen capacity and leadership in research related to respiratory conditions, critical care and sleep medicine in middle and low income countries

Testing a hypothesis

- We want to make inferences about a population from a sample.
  - i.e., we have a hypothesis we want to test

- We need to choose the number of observations to include in a study sample.
  - The larger the sample, the greater the power of the statistical test

=> Sample size determines statistical power

Testing a hypothesis

- What is (statistical) power?

$\Rightarrow$ It is the probability that we will observe an intervention effect (based on data from our sample) when an intervention effect actually exists.

Testing a hypothesis: Example

- Say we want to test a text messaging intervention to remind diabetes patients to take their meds on schedule, which should result in lower HbA1c levels.
  - We randomly assign patients to the **text message intervention** or to the "**usual care" control** groups.
  - Define the mean difference between HbA1c scores in the intervention and non intervention groups as our effect size:

$$\delta_{diff} = \delta_{tx} - \delta_{no\ tx}$$

δ_{diff} is the "effect size"

Testing a hypothesis: Example

- We specify mutually exclusive "null" and "alternative" hypotheses":
  - Null hypothesis: there is no difference in mean HbA1c between patients receiving the text messages and those not:

$$H_0: \delta_{diff} = 0$$

  - Alternative (2-sided) hypothesis : mean HbA1c differs between patients who do and do not get the text intervention:

$$H_A: \delta_{diff} \neq 0$$

The truth vs what we observe

- What is the truth?
  - Either the intervention is effective or it isn't (we don't know which)
- What do we observe (we only do the study in a sample of the intended population)
  - Either we will observe an effect on HbA1c level when we analyze the data from our sample or we won't
- We want to optimize our chances of:
  - Observing an effect in our data **if** the truth is that the intervention really works.
  - NOT observing an effect **if** the truth is that the intervention really doesn't work

Testing a Hypothesis

What we see (in our sample)	The "Truth" (unknown to us) Our text intervention is:	
	Effective	Not Effective
We observe a difference in HbA1c	OK	error!
We don't observe a difference in HbA1c	error!	OK

Testing a hypothesis

- Let's think about these possibilities in terms of probabilities

Testing a Hypothesis

What we see (in our sample)	The "Truth" (unknown to us) Our text intervention: Works	 Doesn't Work
We observe a difference in HbA1c	Probability of deciding **there is an effect** when there really is one	Probability of deciding **there is an effect** when there really isn't one
We don't observe a difference in HbA1c	Probability of deciding **there is no effect**, when there really is one	Probability of deciding **there is no effect** when there really isn't one

Testing a Hypothesis

What we see (in our sample)	The "Truth" (unknown to us) Our text intervention: Works	 Doesn't Work
We observe a difference in HbA1c	Power = $1 - \beta$	Type I error $= \alpha$
We don't observe a difference in HbA1c	Type II error $= \beta$	$1 - \alpha$

Testing a Hypothesis

What we see (in our sample)	The "Truth" (unknown to us) Our text intervention: Works	 Doesn't Work
We observe a difference in HbA1c	Power = $1 - \beta$ Usually $\geq .80$	Type I error $= \alpha$ Usually $<.05$
We don't observe a difference in HbA1c	Type II error $= \beta$ Usually $< .20$	$1 - \alpha$ Usually $\geq .95$

Testing a Hypothesis

- Setting $\alpha = .05$ is traditional (and arbitrary)
  - i.e., we accept a 5/100 (or 1 in 20) chance of making type I error (i.e., deciding that there IS an effect when there really ISN'T)
- Setting power = .80, (i.e. $\beta = .20$) means we expect to have at least an 80% chance of detecting an effect when there really is one.

Testing a Hypothesis

- So… if we want to have 80% power and a 5% significance level how many patients do we need?

Sample Size (N) depends on

- Type I error = α
- Power = 1- β
- Effect size = Estimate of expected treatment effect, $\delta_{diff} = \delta_{tx} - \delta_{no\ tx}$
- Estimate of variability in treatment effect, σ
  - Will therapy affect everyone to the same degree or does the effect vary substantially, affecting some patients a lot and others very little?

Sample Size

- Some common sample size applications assume that your data are normally distributed

A simple sample size calculation:

$$2n = \frac{4(Z_{1-\alpha} + Z_{1-\beta})^2 \ \sigma^2}{\delta^2}$$

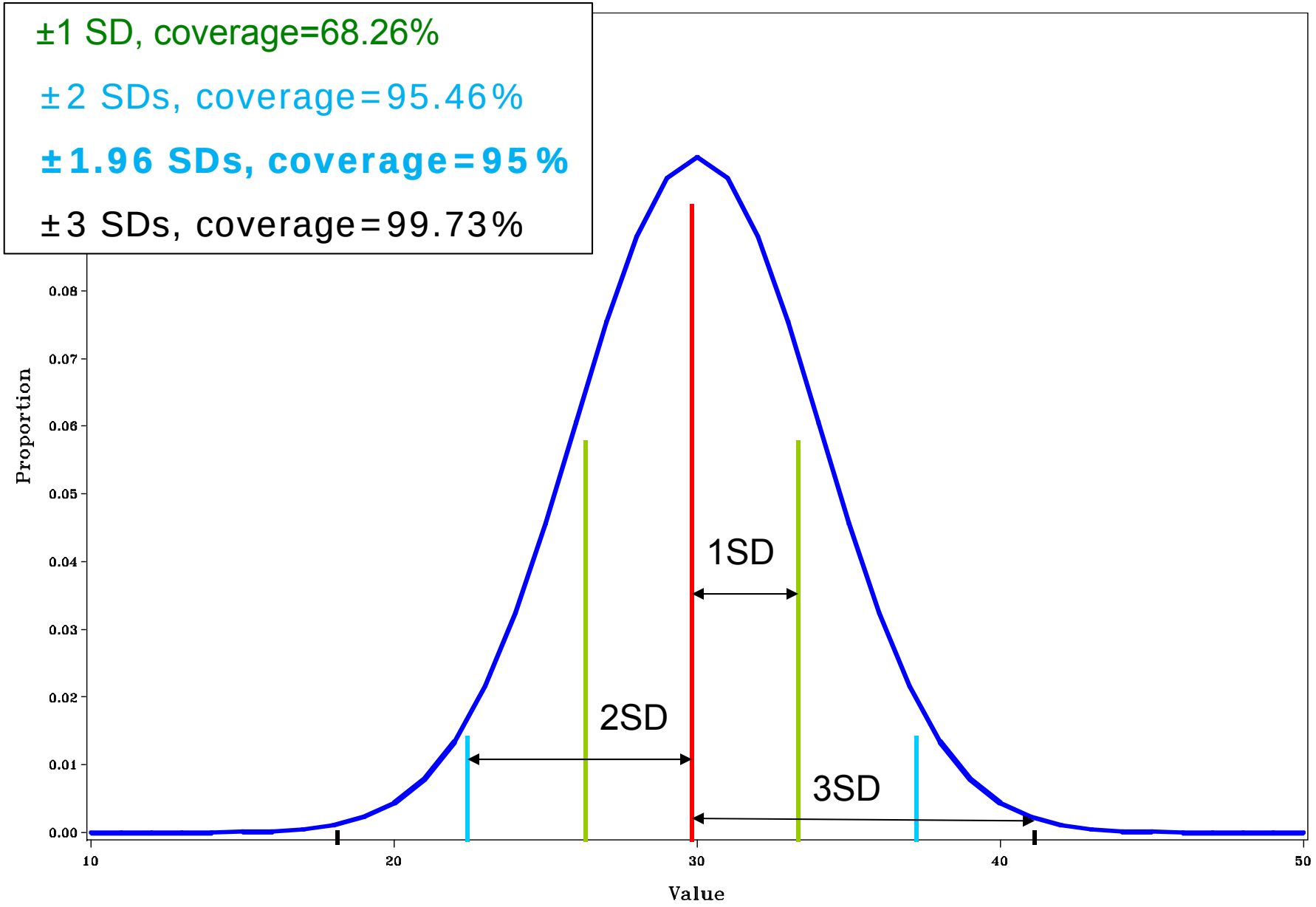
Normal Distribution

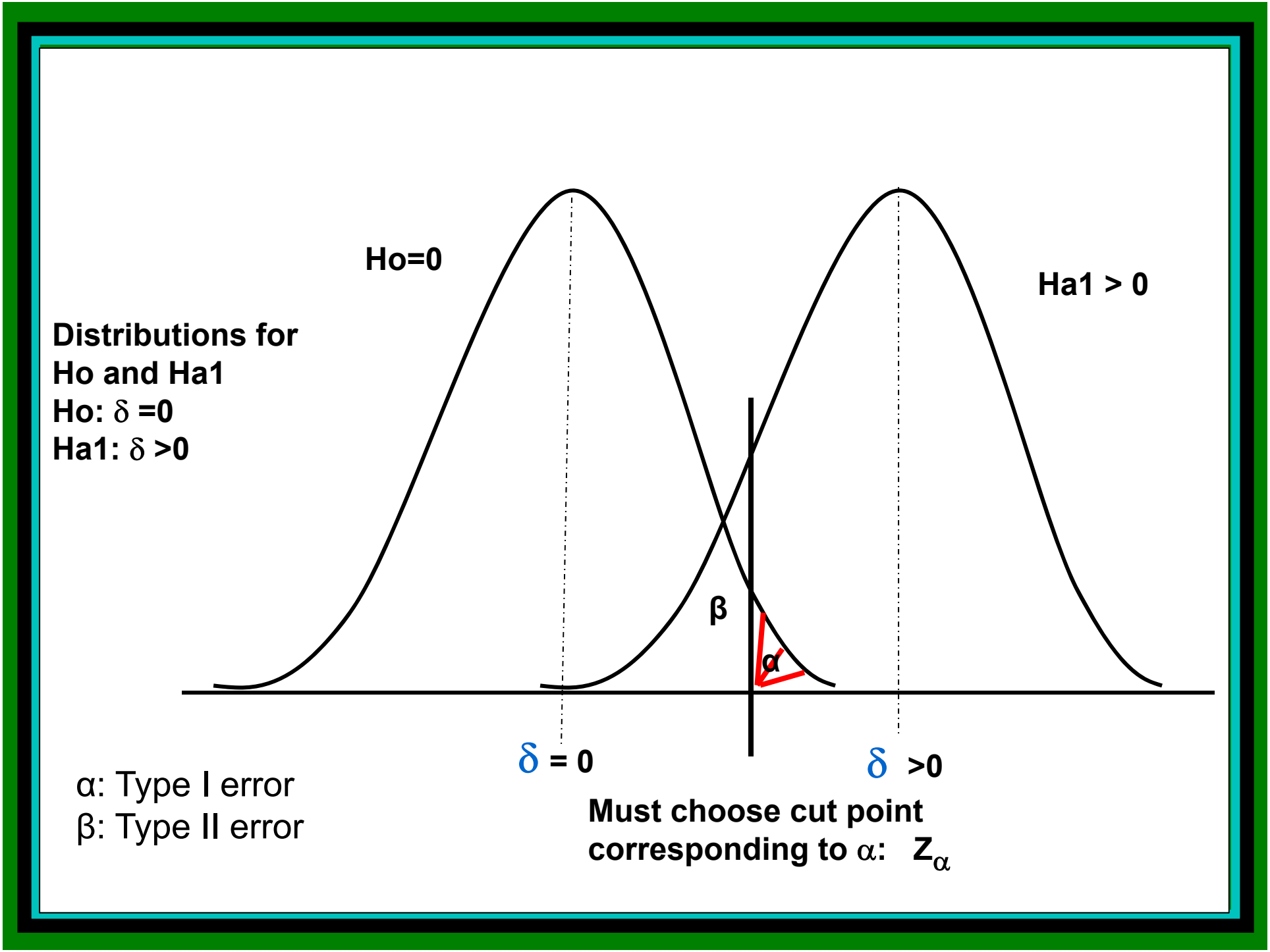
±1 SD, coverage=68.26%

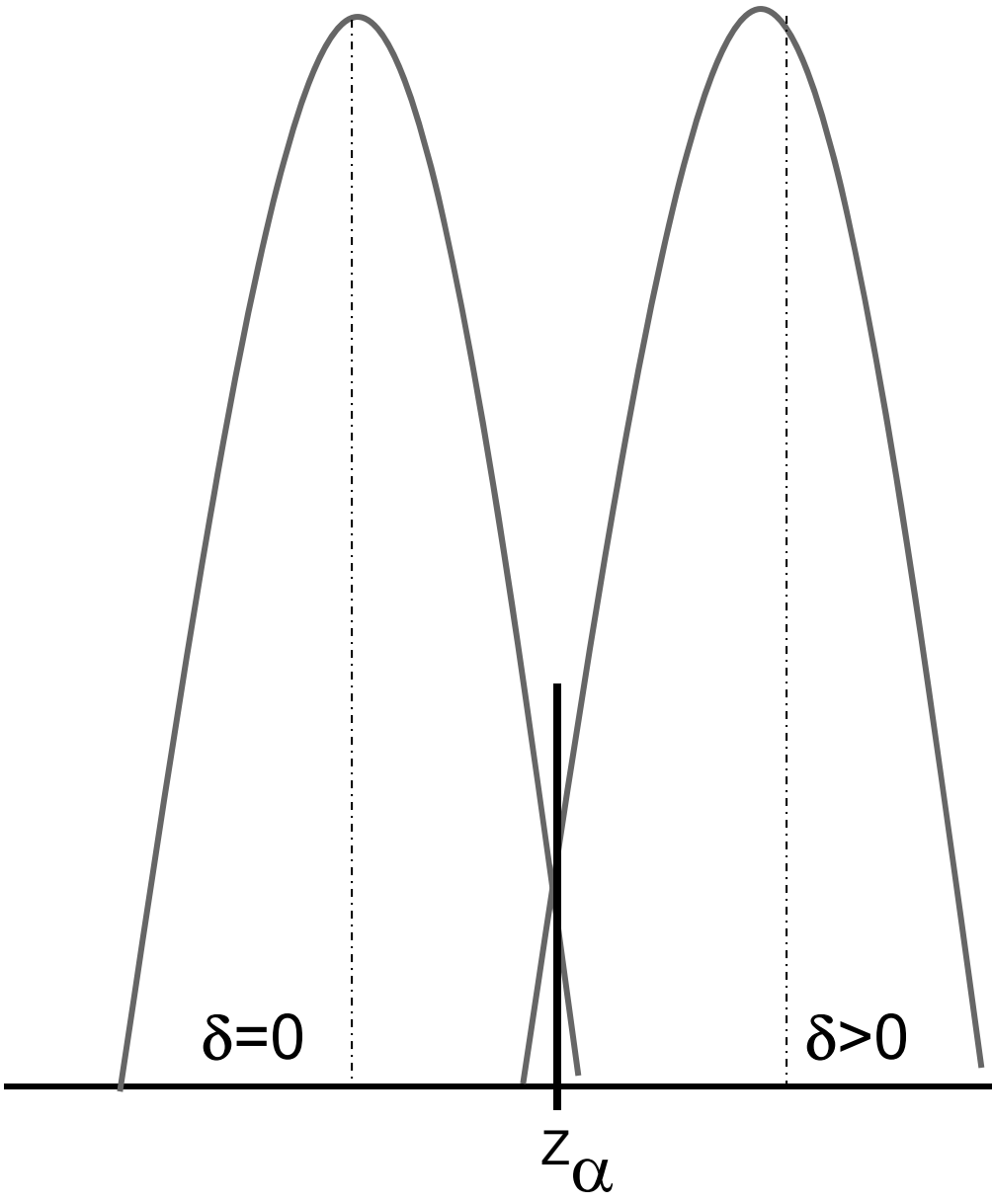
±2 SDs, coverage=95.46%

±1.96 SDs, coverage=95%

±3 SDs, coverage=99.73%




Distributions for
Ho and Ha1
Ho: δ =0
Ha1: δ >0
Ho=0
Ha1 > 0
β
α
δ = 0
δ >0
Must choose cut point
corresponding to α: Z_α
α: Type I error
β: Type II error



Increase in Sample Size
Tightens the Distributions
Decreases the Overlap

Both α and β are decreased

Trade-offs between sample size, power and effect size

- If we want to increase power, say from 80% to 90%, we must increase sample size
- If we want to detect a smaller effect size, say a 10% increase in survival instead of a 20% increase, we must increase sample size

Quiz

What is β?

- Type II error: The probability of failing to reject the null hypothesis when it is false

What is α?

- Type I error: The probability of rejecting the null hypothesis when it is true

What is 1-β?

- Power: the probability of accepting the alternative hypothsis when it is true.

Resources for power and sample size calculations

- http://davidmlane.com/hyperstat/power.html
- http://homepage.stat.uiowa.edu/~rlenth/Power/

Sampling Methods

EnCoRE Presentation, Feb 17, 2015

Sampling

- Methods for selecting subjects for study e.g., people, organizations, trees, etc.) from a population of interest .
- Often we want to generalize results of study to the larger population.

Sampling

- Theoretical population vs accessible population
- Usually not feasible to sample full population.
  => Sample some members of a theoretical population as representative of the entire population
- Multi-step process
  - Identify population of interest
  - Identify accessible population
  - Develop sampling frame
  - Draw sample
  - Contact/recruit subjects/participants

Types of sampling

- Convenience sampling = identify a group of people you can get to "conveniently“
  - Examples: hospital staff, market women, senior housing
  - No "formal" sampling frame is used
  - Not able to calculate confidence intervals
  - Useful for many purposes

- Probability sampling = random selection
  - SRS (‘simple random sample’)
  - Stratified sample
  - CS (‘cluster sample’)

Random Sampling

- Involves creation of a "Sampling frame" (list of members of group to be sampled).
- Assures equal probability of selection for all members of the population.
- Estimates will be "unbiased" (precise statistical meaning: average of multiple samples will have population mean)
- Note: random does not mean "haphazard"

Simple Random Sampling (SRS)

- SRS Process:
  - Develop sampling frame: List accessible population of N subjects from which n subjects will be drawn (e.g., all individuals/HHs).
  - Use random process, e.g. random number table, to generate "n" numbers between 1 and N.
  - Identify "n" individuals in sample corresponding to the "n" numbers generated.

- Examples: phonebook - random digit dialing.

Simple Random Sampling (SRS)

- Advantages:
  - Most basic form of sampling
  - The gold standard to which other methods are compared.
- Disadvantages:
  - All individuals/HH must be identified prior to sampling
  - May be unrealistic - time, money.
  - Selected individuals/HHs may be highly dispersed. Visiting each may be very time consuming

Stratified Random Sampling

- Divide population in homogeneous (mutually exclusive) subgroups (i.e., “strata”) and do SRS within each group
- Advantages
  - Enough cases from each strata to make meaningful inferences on key subgroups (e.g., small minority groups).
  - Generally have more statistical precision than SRS **if** the strata are homogeneous:

    => the variability within-groups is lower than the variability for the population as a whole.

    => smaller sample sizes
- Examples of common strata: age, race, gender, educational attainment level, socioeconomic status, rare traits, diseases or conditions

Cluster Sampling (CS)

- CS Process
  - Construct sampling frame using groups or clusters of individuals (or HHs) without identifying or listing each one.
    - Cluster = villages, towns, districts, urban blocks, etc.
  - List clusters
  - Take a random sample of the clusters
  - Obtain list of individuals/HHs only for those clusters selected in the sample
  - Sample a random sample of individuals/HHs within each selected cluster

Cluster Sampling (CS)

- Advantages:
  - Economy: listing costs/travel costs.
  - Feasibility: most countries will have lists of population by groups (villages, towns)
  - Allows a larger sample size than SRS due to decreased costs
    - for same resources (money, person-time), it's possible to gain greater precision with cluster sampling
- Disadvantages:
  - Estimate not as precise as SRS for the same sample size.
    - Design Effect = Variance Cluster Sampling/Variance SRS. Factor by which to increase CS sample size to obtain same precision as for SRS
- Example: WHO Expanded Programme on Immunization (EPI) sampling method.

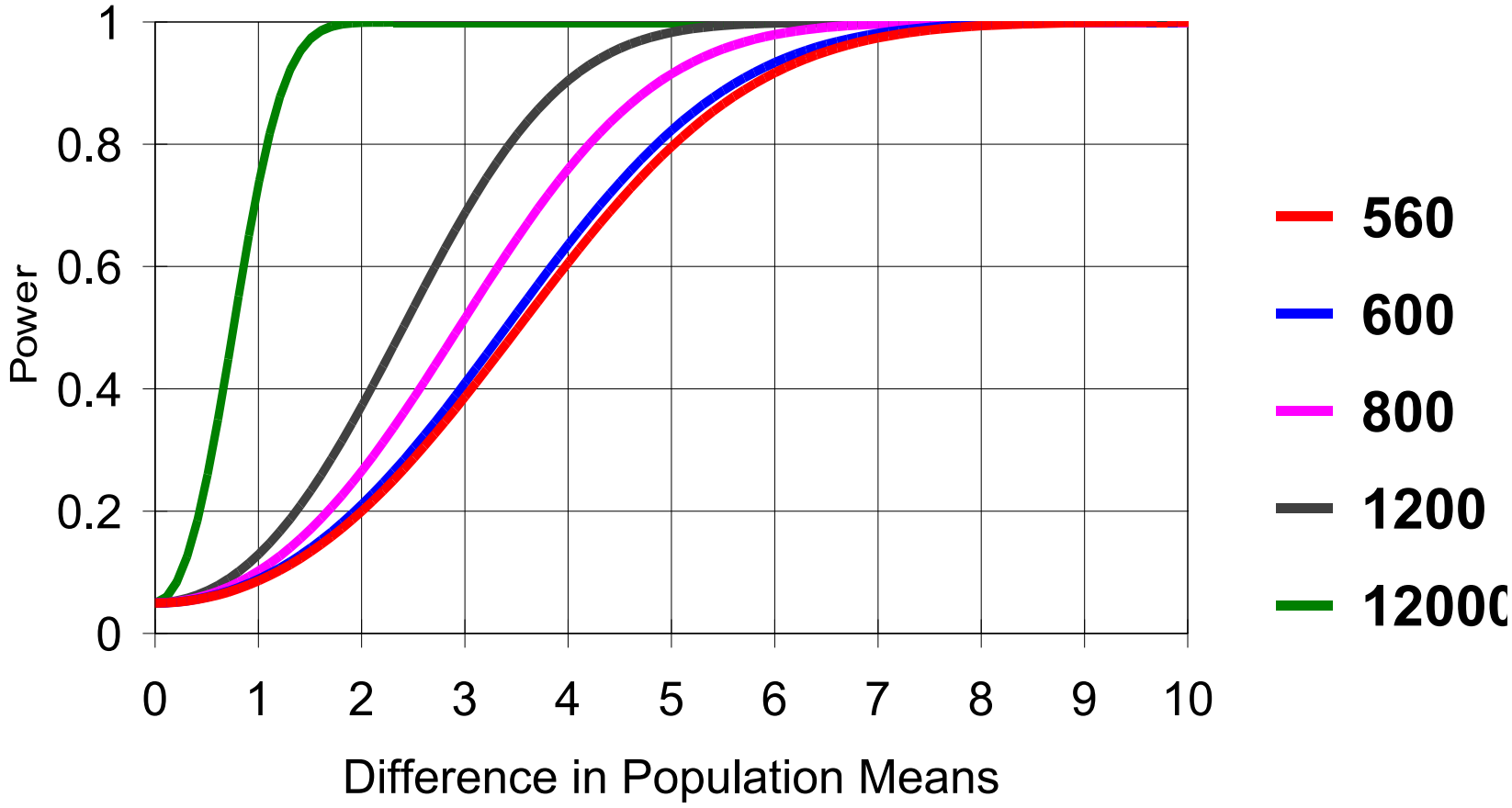
Cluster Sampling (CS)

- Considerations
  - Budget – transport costs can be high, esp. in rural areas
  - Travel/logistics
    - Define center/periphery of urban areas.
    - How to account for appartments vs houses – need to establish unambiguous methods, eliminate personal choice of researchers
    - Number of surveys in given area per day.
  - Supplies
  - Technical
    - Can estimate optimal cluster size if you know:
      - » transport costs to each cluster
      - » cost for interviewing each respondent
      - » Roh

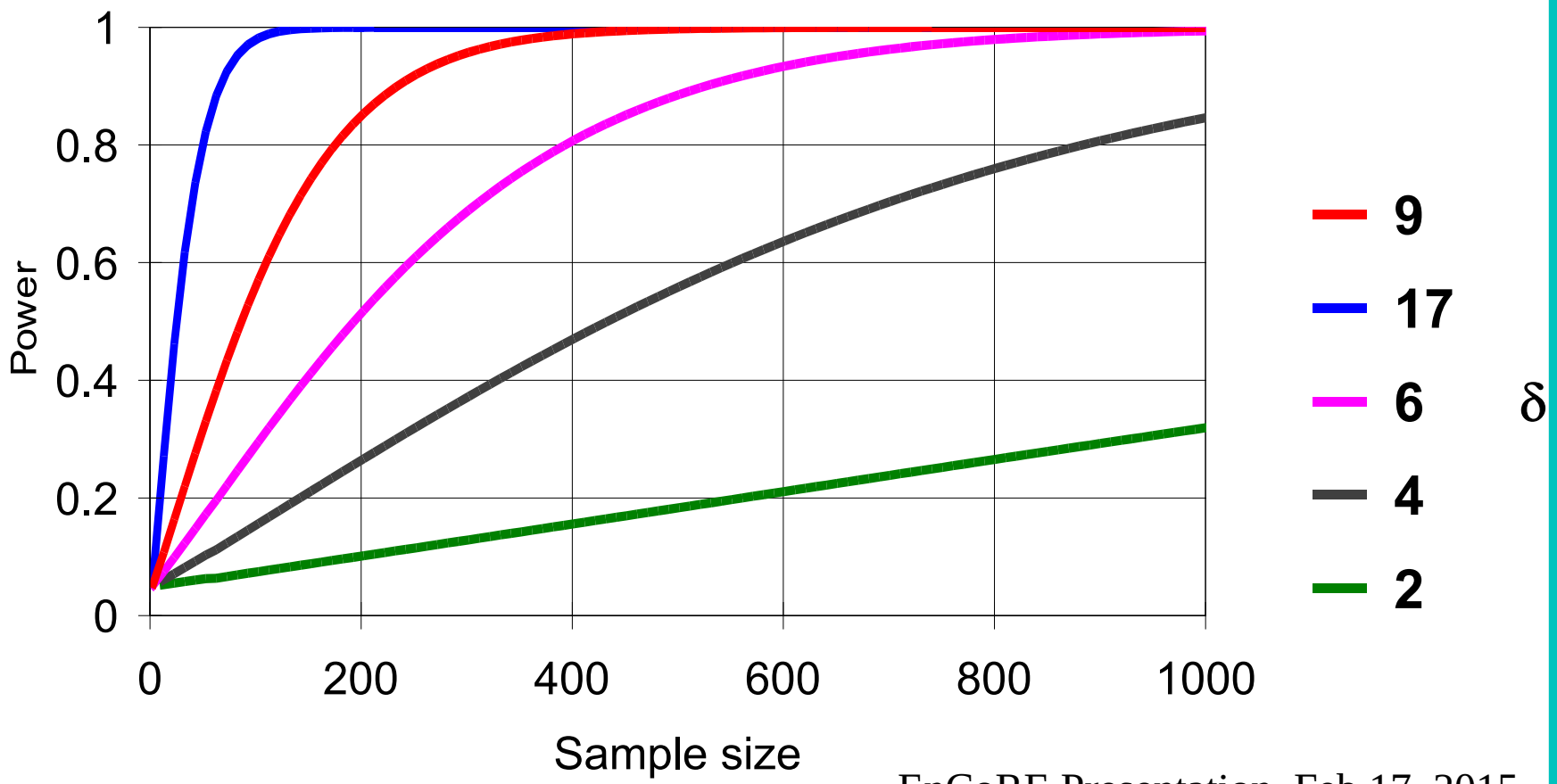
Questions?

Thank You!

How sample size affects the power to detect differences



How δ affects the relationship between power and sample size



How variability affects the relationship between power and sample size