Clinical and Translational Science Institute / CTSI

at the University of California, San Francisco

Sample Size and Power

Steven R. Cummings, MD

San Francisco Coordinating Center

CPMC Research Institute

The Secret of Long Life

- Resveratrol
- In the skin of red grapes
- Makes mice
  - Run faster
  - Live longer

Mimics 'sirtuin', Senses energy and controls DNA transcription; produces more mitochondria.

What I Want to Show

Consuming resveratrol prolongs healthy life

What I Need to Know: Ingredients for Estimating Sample Size

Testable Hypothesis

Type of Study

Statistical Test: Type of Variables

Effect Size

Power and Alpha

My Research Question

Does consuming resveratrol lead to a long and healthy life?

Testable Hypothesis

What's Wrong?

Does consuming resveratrol lead to a long and healthy life?

- **Vague**
- **To estimate sample size, it must be a testable hypothesis measurable**

“Consuming Resveratrol”

I could measure resveratrol intake in a cohort, then follow-up for survival free of disease …

"Consuming Resveratrol"

- Most rigorous design: randomized placebo-controlled trial
- Comparing red wine to placebo would be difficult
- But resveratrol supplements are widely available

A Specific Predictor

- “Consuming resveratrol”
- ***Resveratrol supplements vs. placebo***

I Need a Measureable Outcome

- “Prolongs healthy life”
- Lower mortality rate

Measurable (Specific) Outcome

- "Consuming resveratrol" =
  - taking resveratrol supplements vs. taking placebo
- "Prolong ~~healthy~~ life" =
  - reduces all-cause mortality

Do ***people*** randomized to get a resveratrol supplement have a lower mortality rate than those who get a placebo?

The Question is Becoming a Hypothesis

Do people randomized to get a resveratrol supplement have a lower mortality rate than those who get a placebo?

Testable Hypothesis

I Must Specify the Population

Do people randomized to get a resveratrol supplement have a lower mortality rate than those who get a placebo?

In Whom?

Elderly men and women

(≥ 70 years)

In Whom?

- Elderly men and women (≥ 70 years)
  - Because the question applies to both genders
  - Because they have a relatively high mortality rate
    - I'll have enough 'events' within a feasible period of study

The Research Hypothesis (AKA the 'alternative' hypothesis)

Men and women > age 70 years randomized to get a resveratrol supplement have a lower mortality rate than those who get a placebo.

However, it cannot be tested statistically

Statistical tests only reject <u>null</u> hypothesis – that there is <u>no</u> effect

The Null Hypothesis

Men and women > age 70 years randomized to get a resveratrol supplement ***do not*** have a lower mortality rate than those who get a placebo.

- Can be rejected by statistical tests

Ingredients for Sample Size

- Testable Hypothesis
- **Type of Study: Descriptive or Analytical**

Descriptive Studies

- Only one variable
- Sample size is based on the desired width of the confidence interval
- What proportion of centenarians (≥100 years old) take resveratrol supplements?
  - Confidence interval for proportions
- What is the mean red wine intake of centenarians?
  - Confidence interval for the mean

Analytical Studies

- Predictor and outcome variable(s)
- Comparisons
  - For example:
    - Compare the mean red wine intake in centenarians ***vs***. 60-80 year olds
    - People who get resveratrol have lower mortality ***than*** those who get placebo

Ingredients for Sample Size

- Testable Hypothesis
- Type of Study: Descriptive or Analytical
- **Statistical test**
  - **Type of variables**

Type of Statistical Tests Depends on the Types of Variables

Predictor, outcome	Dichotomous	Continuous
Dichotomous	**Chi-squared**	**T test**
Continuous	**T test**	**Correlation**

This works for most study planning

Statistical Test

The Types of Variables?

Men and women > age 70 years randomized to receive a resveratrol supplement do not have lower mortality rate than those who receive placebo

- **Dichotomous:** resveratrol or placebo
- **Continuous**: mortality rate

What's wrong?

The Types of Variables?

Men and women > age 70 years randomized to receive a resveratrol supplement do not have lower mortality rate than those who receive placebo

- **Dichotomous:** resveratrol or placebo
- ~~**Continuous**~~: **mortality *rate***
  - ***It is a proportion at certain times***
  - ***For example, 3% at 1 year***

Type of Study

Type of Statistical Tests Depends on the Types of Variables

Predictor, outcome	Dichotomous	Continuous
Dichotomous	**Chi-squared**	**T test**
Continuous	**T test**	**Correlation**

Ingredients for Sample Size

- Testable Hypothesis
- Type of Study: Descriptive or Analytical
- Statistical test
  - Type of variables
- **Effect size**

Estimating the Effect Size

For randomized trials;

- Start with the expected rate in the placebo
- Usually available from population or cohort studies
- In this case, we know the mortality rates by age:
- 3-4% per year*; for a 3 year study: 10%

* ~ mean annual female/males @ 78 yrs

Effect Size

The Hardest Part

What should I assume for the effect of resveratrol on mortality?

Effect Size
The Hardest Part

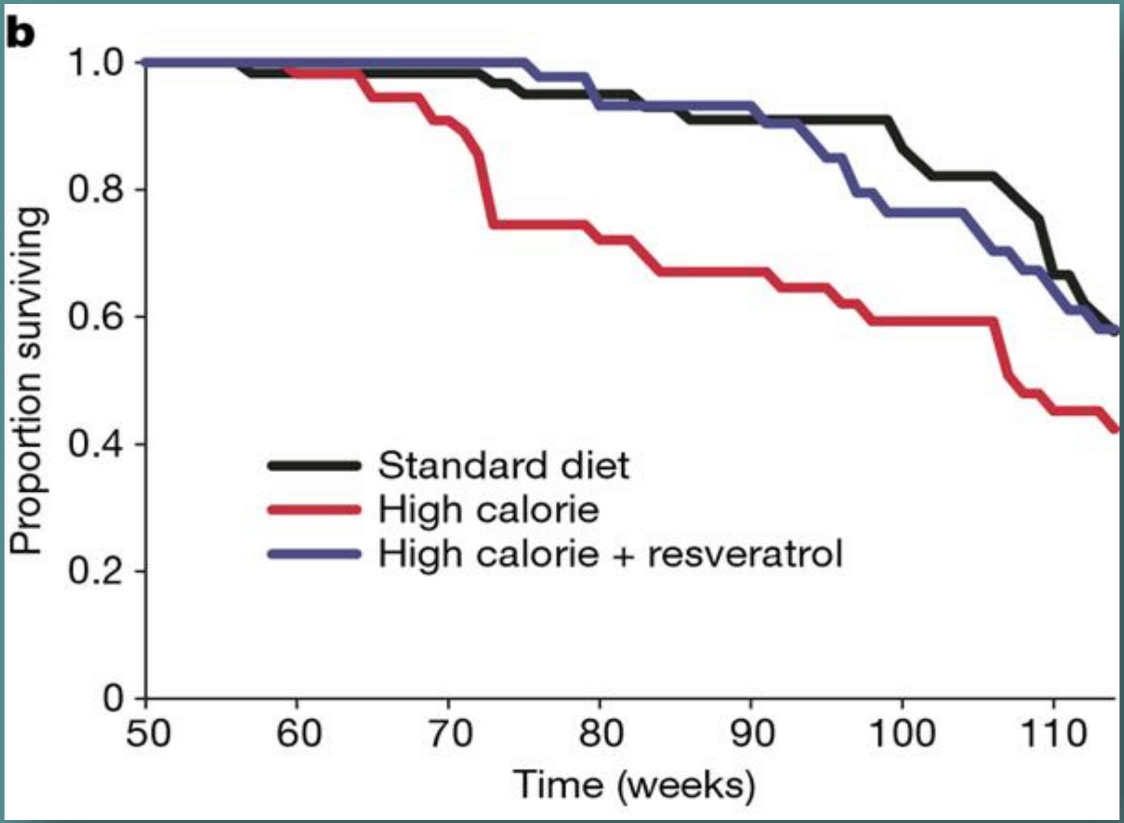
Ways to choose an effect size:

- What is likely, based on other data?
- Do a pilot study
- Estimate based on effect on biomarkers
- What difference is important to detect?
  - "We don't want to miss a _%_ difference"
- What can we afford?

The effect of resveratrol on mortality rate?

- What is likely, based on other data?
  - No data in humans!

Resveratrol pronged survival of mice fed high calorie diet



Baur, Nature 2006

The effect of resveratrol on mortality rate?

- What is likely, based on other data?
- **Pilot study? Small pilot studies generally produce unstable estimates of effects and variance**

Effect Size

The effect of resveratrol on mortality rate?

- What is likely, based on other data?
- Do a pilot study
- Estimate based on effect on biomarkers
  - Biomarkers for mortality and effect of resveratrol

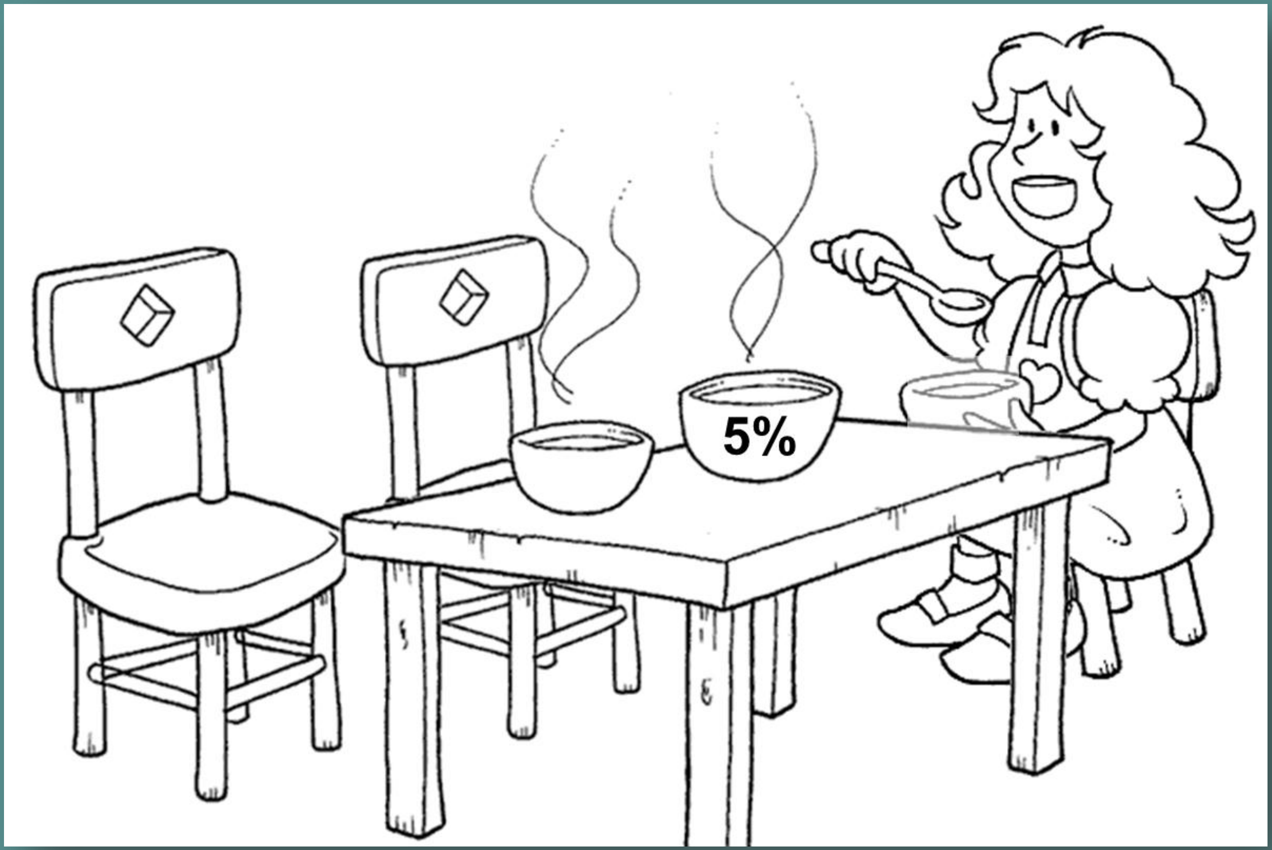
The effect of resveratrol on mortality rate?

- What is likely, based on other data?
- Do a pilot study
- Estimate based on effect on biomarkers

- What difference is important to detect?
  - I don't want to miss a 1% decrease!
- What can we afford?
  - 1%: the trial will be too big & expensive
  - 5%: the study will be smaller and cheaper

Effect Size

The Science of Effect Sizes
Too Large!



The effect of resveratrol on mortality rate?

- What is likely, based on other data?
- **Pilot study??**
  - **Small pilot studies generally produce unstable estimates of effects and variance**

The effect of resveratrol on mortality rate?

- What is likely, based on other data?
- Do a pilot study
- **Estimate based on effect on biomarkers**
  - **Biomarkers for mortality and effect of resveratrol?**

Effect Size

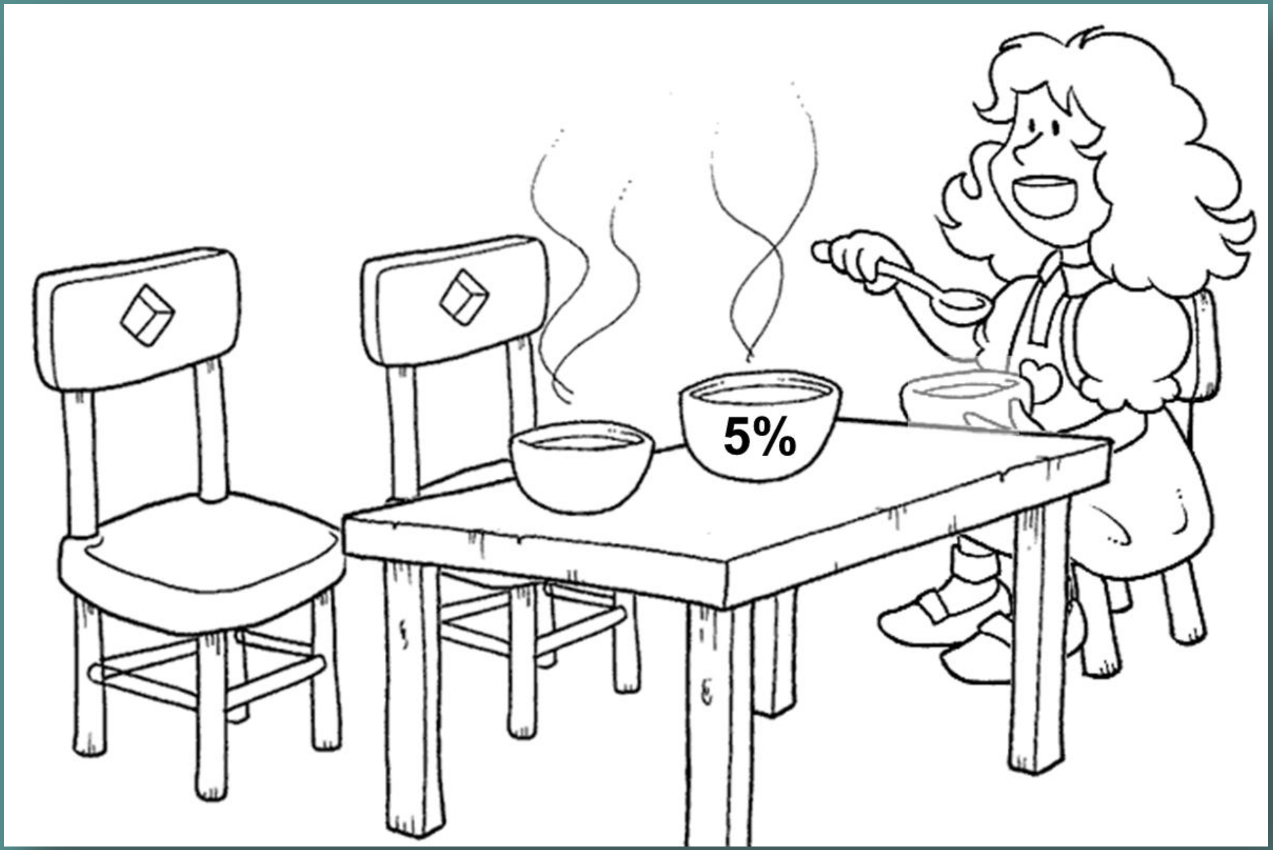
The effect of resveratrol on mortality rate?

- What is likely, based on other data?
- Do a pilot study
- Estimate based on effect on biomarkers

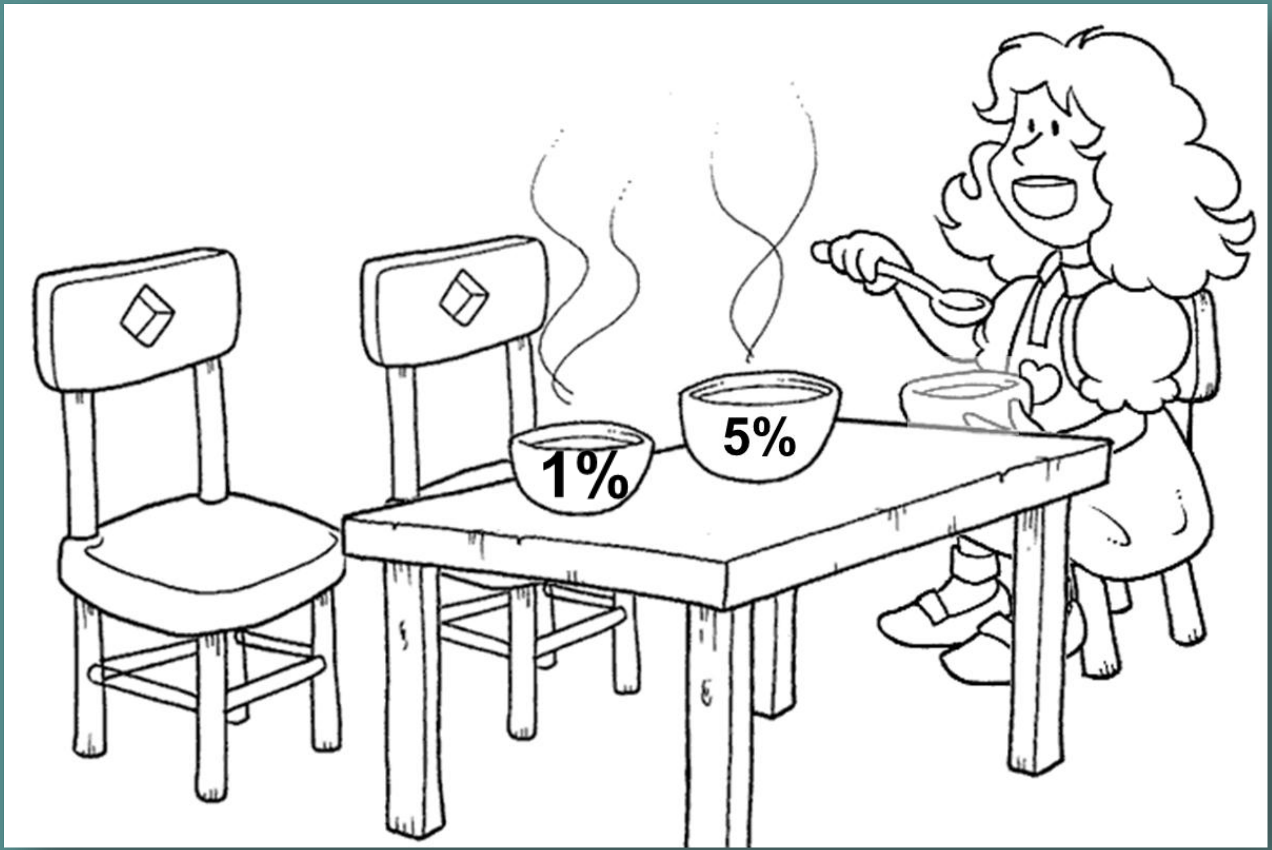
- **What difference is important to detect?**
  - **I don't want to miss a 1% decrease!**
- **What can we afford?**
  - **1%: the trial will be too big & expensive**
  - **5%: the study will be smaller and cheaper**

Effect Size

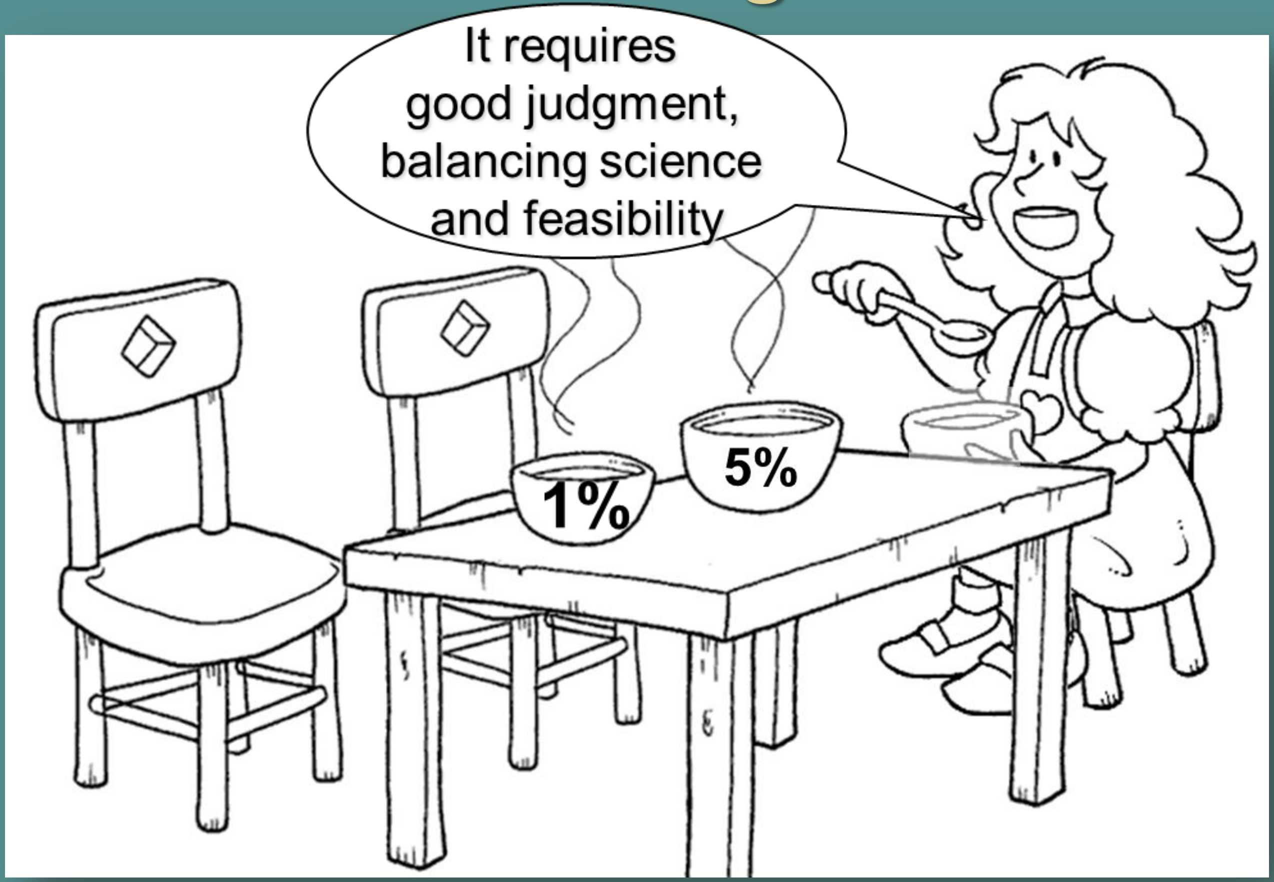
The Science of Effect Sizes
Too Large!



The Science of Effect Sizes
Too Large!



The Science of Effect Sizes
Too Large!



Effect Size

Men and women > age 70 years randomized to receive a resveratrol supplement do not have lower mortality rate than those who receive placebo

- It would be important to find (I don't want to miss) a 20% decrease
- Placebo rate: 10%
- Resveratrol rate: 8%

Ingredients for Sample Size

- Testable Hypothesis
- Type of Study: Descriptive or Analytical
- Statistical test
  - Type of variables
- Effect size
- **Power and alpha**

α (alpha)

The probability of finding a 'significant' result if nothing is going on

"Type 1 error"

To Convince People that an Effect is Not Due to Chance

- Customarily, a result is ‘statistically significant’ if $P<0.05$

In other words,

- Probability of a type I error = 5%
- α (alpha) = 0.05

I Will Need to Convince Skeptics

- *Very* small chance that a positive result is an error

α (alpha) = 0.01

$P<0.01$

- A smaller α means larger sample size

Two-sided vs. one-sided α

- A 2-sided α assumes that the result could go either way
  - Recognizes that you have two chances of finding something that isn't really there
    - Resveratrol decreases mortality
    - Resveratrol increases mortality
- A 1-sided hypothesis assumes that the result could, plausibly, go only one way

One-sided α

- You may believe that your effect could only go one way!
  - Resveratrol is 'natural.' It could not increase mortality!
- Be humble.
  - The history of research is filled with results that contradicted expectations

One-sided α?

- There is a universal consensus: vitamin D is good for muscle and bone!
- Vitamin D trial (Sanders et al, JAMA 2010
  - ~1500 IU of vitamin D/d
  - Significantly increased the risk of falls and fractures in elderly women and men
- A 1-sided test is almost never the best choice

β (beta)

The probability of *missing* this effect size in this sample, if it is really true in the populations

AKA Type 2 error

Power (1 - β)

The probability of *finding* this effect size in this sample, if it is really true in the population

If it's True, I Don't Want to Miss It

The chance of missing the effect (β) is "customarily" 20%

In other words

- Probability of a type II error = 0.20
- β (beta) = 0.20
- **Power** = 1- β : 0.80
- "80% power"

I *REALLY* Don’t Want to Miss It

$\beta = .10$

Power $(1- \beta) = 0.90$

Greater power requires a larger sample size

We Have All the Ingredients

Testable Hypothesis

Type of Study: Analytical (RCT)

Statistical Test: Chi-squared

Effect Size: 10% vs 8%

Power: 0.90; Alpha: 0.05

From Table 6B.2
Comparing two proportions

Upper number: $\alpha = 0.05$ (one-sided) or $\alpha = 0.10$ (two-sided); $\beta = 0.20$
Middle number: $\alpha = 0.025$ (one-sided) or $\alpha = 0.05$ (two-sided); $\beta = 0.20$
Lower number: $\alpha = 0.025$ (one-sided) or $\alpha = 0.05$ (two-sided); $\beta = 0.10$

Smaller of P_1 and P_2	Expected Difference Between P_1 and P_2									
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0.06	7,703	2,107	1,016	615	422	312	243	197	163	139
	9,726	2,650	1,272	769	526	388	301	243	202	171
	12,973	3,518	1,684	1,014	691	508	395	318	263	223
0.07	8,765	2,369	1,131	680	463	340	263	212	175	148
	11,076	2,983	1,419	850	577	423	327	263	217	183
	14,780	3,965	1,880	1,123	760	555	429	343	283	239
0.08	9,803	2,627	1,244	743	502	367	282	227	187	158
	12,393	3,308	1,562	930	627	457	352	282	232	195
	16,546	4,401	2,072	1,229	827	602	463	369	303	255
0.09	10,816	2,877	1,354	804	541	393	302	241	198	167
	13,679	3,626	1,702	1,007	676	491	377	300	246	207
	18,270	4,827	2,259	1,333	893	647	495	393	322	270
0.10	11,804	3,121	1,461	863	578	419	320	255	209	175
	14,933	3,936	1,838	1,083	724	523	401	318	260	218
	19,952	5,242	2,441	1,434	957	690	527	417	341	285

From Table 6B.2

- Sample size: 4,401
- Per group
- Total: 8,802
- Does not include drop-outs
  - 20% drop-out: 11,002 total sample size!

Appropriate Responses

- Shock and awe
- Depression
- Anxiety
- Consider alternative approaches

Alternatives

- Tweak α: one-sided
  - Almost never appropriate
- Tweak the power: 0.80

From Table 6B.2
Comparing two proportions

Upper number: $\alpha = 0.05$ (one-sided) or $\alpha = 0.10$ (two-sided); $\beta = 0.20$
Middle number: $\alpha = 0.025$ (one-sided) or $\alpha = 0.05$ (two-sided); $\beta = 0.20$
Lower number: $\alpha = 0.025$ (one-sided) or $\alpha = 0.05$ (two-sided); $\beta = 0.10$

Smaller of P_1 and P_2	Expected Difference Between P_1 and P_2									
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0.06	7,703	2,107	1,016	615	422	312	243	197	163	139
	9,726	2,650	1,272	769	526	388	301	243	202	171
	12,973	3,518	1,684	1,014	691	508	395	318	263	223
0.07	8,765	2,369	1,131	680	463	340	263	212	175	148
	11,076	2,983	1,419	850	577	423	327	263	217	183
	14,780	3,965	1,880	1,123	760	555	429	343	283	239
0.08	9,803	2,627	1,244	743	502	367	282	227	187	158
	12,393	3,308	1,562	930	627	457	352	282	232	195
	16,546	4,401	2,072	1,229	827	602	463	369	303	255
0.09	10,816	2,877	1,354	804	541	393	302	241	198	167
	13,679	3,626	1,702	1,007	676	491	377	300	246	207
	18,270	4,827	2,259	1,333	893	647	495	393	322	270
0.10	11,804	3,121	1,461	863	578	419	320	255	209	175
	14,933	3,936	1,838	1,083	724	523	401	318	260	218
	19,952	5,242	2,441	1,434	957	690	527	417	341	285

Alternatives

- Tweak α: one-sided
  - Almost never appropriate
- Tweak the power: 0.80
- Modest effect: 3,308 (6,616 total)

Alternatives

- Increase the effective size
  - 10% vs. 6%

From Table 6B.2
Comparing two proportions

Upper number: $\alpha = 0.05$ (one-sided) or $\alpha = 0.10$ (two-sided); $\beta = 0.20$
Middle number: $\alpha = 0.025$ (one-sided) or $\alpha = 0.05$ (two-sided); $\beta = 0.20$
Lower number: $\alpha = 0.025$ (one-sided) or $\alpha = 0.05$ (two-sided); $\beta = 0.10$

Smaller of P_1 and P_2	Expected Difference Between P_1 and P_2									
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0.06	7,703	2,107	1,016	615	422	312	243	197	163	139
	9,726	2,650	1,272	769	526	388	301	243	202	171
	12,973	3,518	1,684	1,014	691	508	395	318	263	223
0.07	8,765	2,369	1,131	680	463	340	263	212	175	148
	11,076	2,983	1,419	850	577	423	327	263	217	183
	14,780	3,965	1,880	1,123	760	555	429	343	283	239
0.08	9,803	2,627	1,244	743	502	367	282	227	187	158
	12,393	3,308	1,562	930	627	457	352	282	232	195
	16,546	4,401	2,072	1,229	827	602	463	369	303	255
0.09	10,816	2,877	1,354	804	541	393	302	241	198	167
	13,679	3,626	1,702	1,007	676	491	377	300	246	207
	18,270	4,827	2,259	1,333	893	647	495	393	322	270
0.10	11,804	3,121	1,461	863	578	419	320	255	209	175
	14,933	3,936	1,838	1,083	724	523	401	318	260	218
	19,952	5,242	2,441	1,434	957	690	527	417	341	285

Increasing the Effect Size

- 10% vs. 6%
- Makes a **big** difference!
- 769 / group; 1,538 total (no dropouts)
- However, still large (and not affordable)
- Not believable

Alternatives: A New Hypothesis

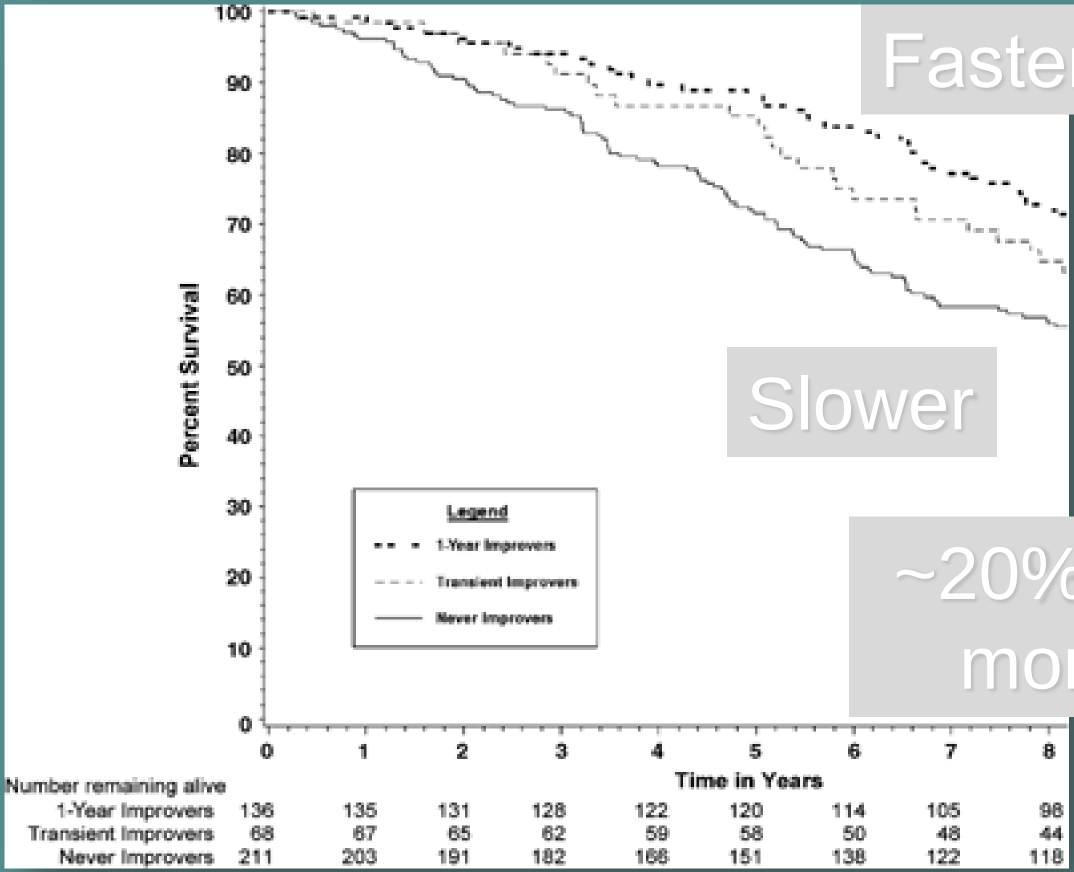
- Change the outcome measure
  - Continuous measurement
  - A precise measurement
- A ‘surrogate’ for mortality rate
  - Strongly associated with mortality rate

    --and--
  - Likely to be influenced by resveratrol
- Walking speed

Mice on Resveratrol

- **Mice fed resveratrol**
  - Live 25% longer
  - Run faster
  - Have greater endurance

Increased gait speed (0.1 m/s) in 1 year and survival over 8 years

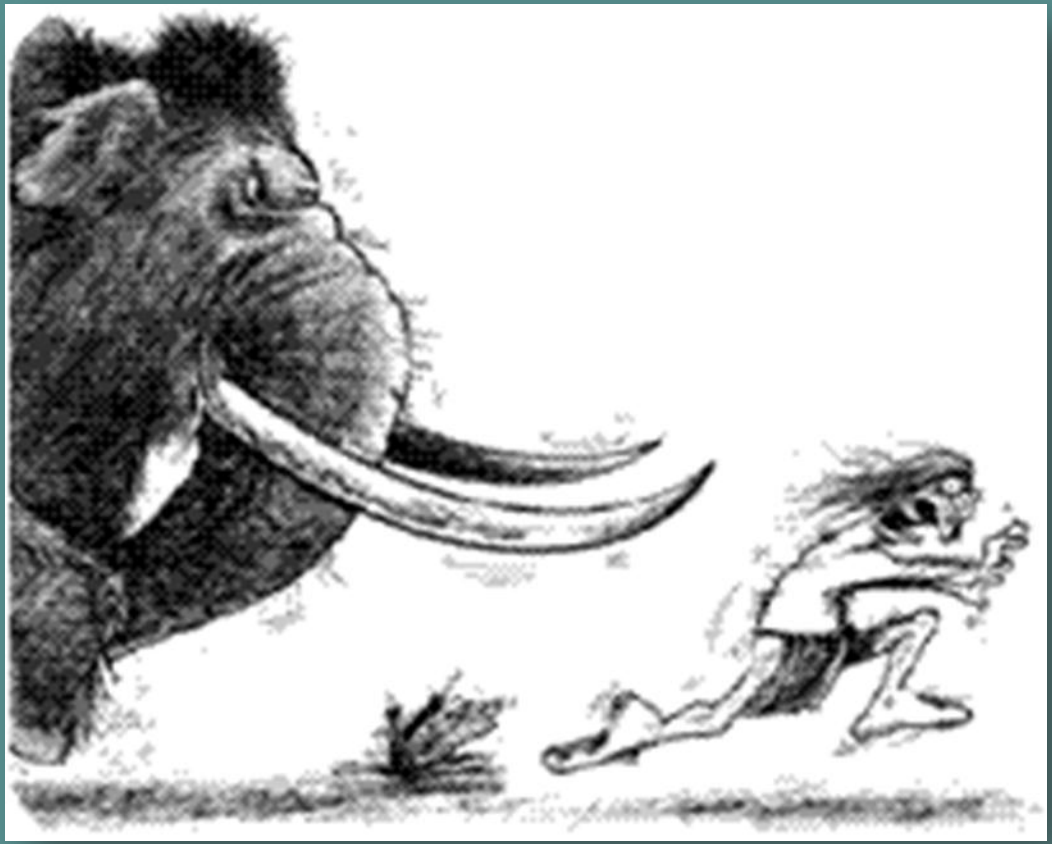


Faster by ≥0.1 m/s

Slower

~20% decreased mortality rate

Fast Walkers have Lower Mortality
An Evolutionary Perspective



Walking Speed

- Easy and cheap!
- Measure time over a 6 meter course

New Endpoint Needs a New Hypothesis

Among 70 year old men and women, those randomized to resveratrol will not have a greater change in walking speed than those assigned to placebo.

The New Ingredients

- New testable hypothesis
- Type of study: analytical RCT
- Statistical test: **?**
  - Continuous variable (walking speed)
  - Difference between means (pbo vs. tx)

Type of Statistical Tests Depends on the Types of Variables

Predictor, outcome	Dichotomous	Continuous
Dichotomous	**Chi-squared**	**T test**
Continuous	**T test**	**Correlation**

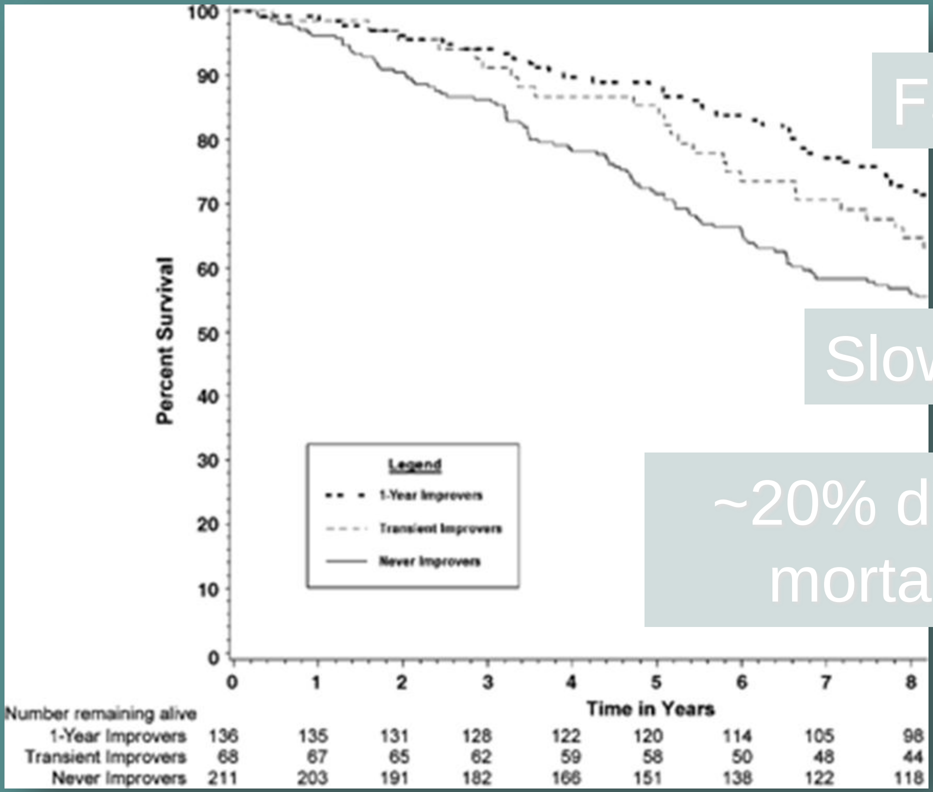
The New Ingredients

- New testable hypothesis
- Type of study: RCT
- Statistical test: **t-test**
- >Effect size: "standardized effect size" **E/S**

Standardized Effect Size: E/S

- Sample size depends on the ratio of E/S
  - E: Effect size (difference in means)
  - S: Standard deviation for measurement
- You need smaller sample size if
  - Greater effect (E)
  - More precise measurement (lower (SD)

Effect Size for Change in Walk Speed?



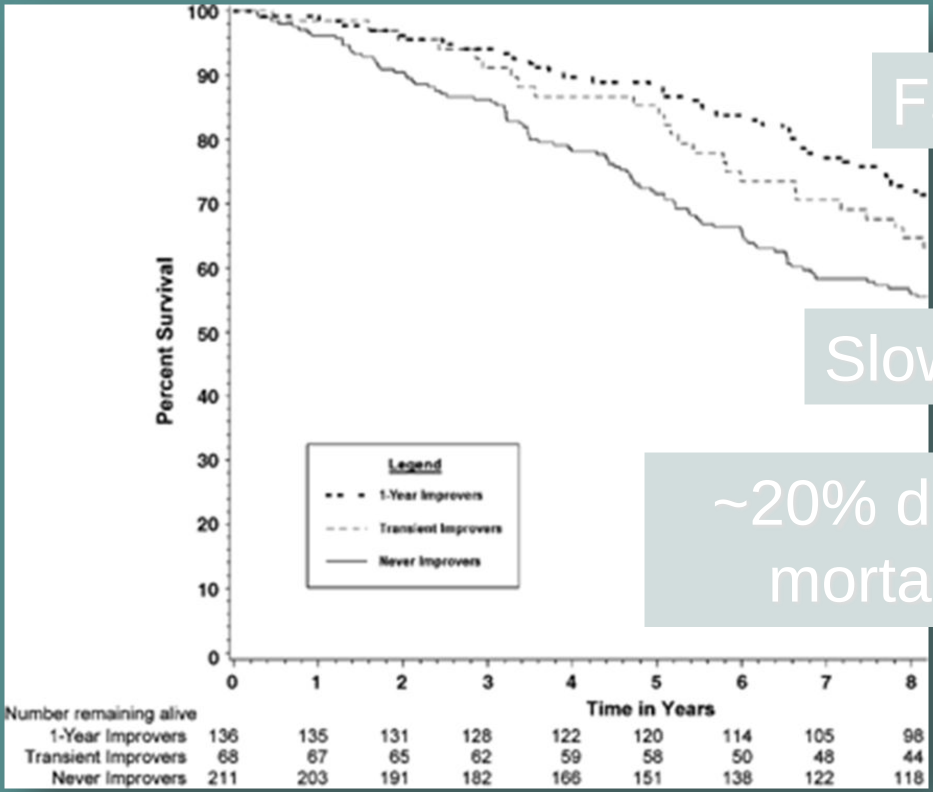
Faster by ≥0.1 m/s

Slower

~20% decreased mortality rate

Human Cohort Studies

0.1 m/s corresponds to a 20% decreased mortality rate



Faster by ≥0.1 m/s

Slower

~20% decreased mortality rate

What we need to determine E/S for our RCT

- Effect size (E) for change in walking speed
  - Decline in the placebo group = -0.1 m/s
  - Prevented by resveratrol = 0 m/s
  - Difference (effect size) = 0.1 m/s

What we need to determine E/S for our Trial

- Effect size (E) for change in walking speed

  +0.1 m/s

- I need **S**: Standard deviation
- Standard deviation of the ***change***

S (Standard Deviation of Change)

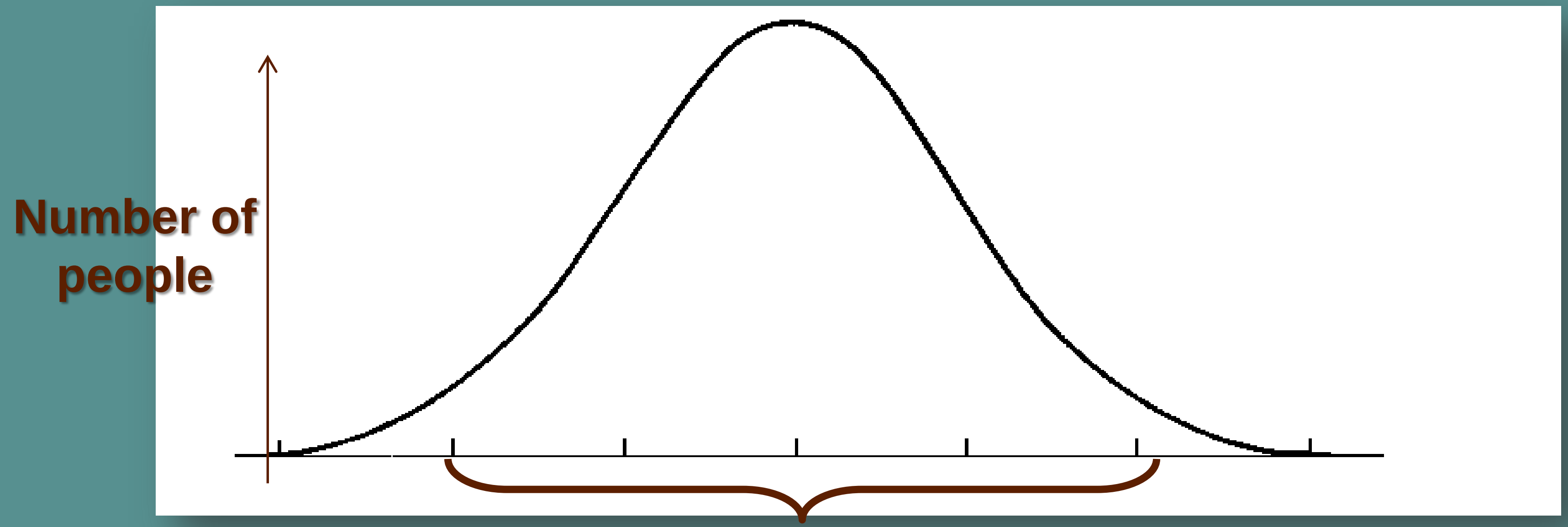
- Standard deviation for the measurement
  - Cross-sectional data: 0.25 m / sec
- However, we are interested in *change*
- Standard deviation of *change* in speed?
  - Often more difficult to find because cross-sectional surveys are more common than longitudinal studies

What if you don't know the SD?

- Standard deviation of *change* in speed?
- If you cannot find data from other studies
- Alternatives
  - Pilot study: measure change in 3 or 6 mo.s
  - Or, a well educated guess

Estimating an S.D.
the 1/4 rule

~ 4 S.D.s across a 'usual' range of values

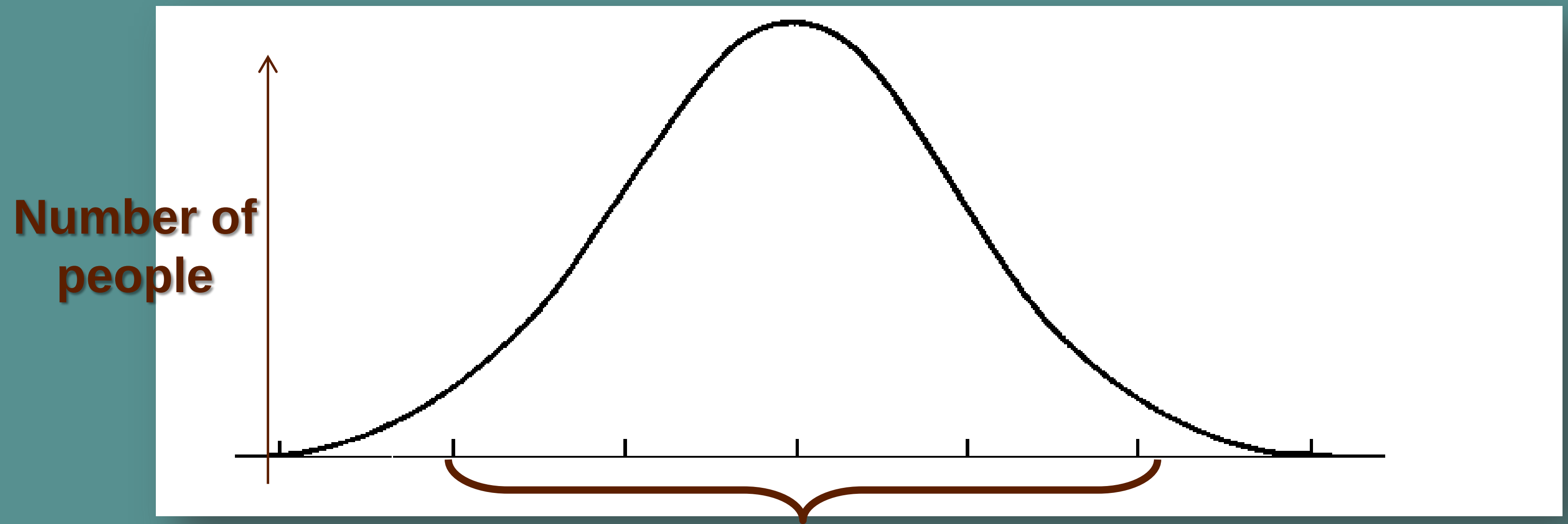


So, 1 S.D. will = 1/4 of the range

Estimating an S.D.

the 1/4 rule

~ 4 S.D.s across a ‘usual’ range of values



Change in Walking Speed

Estimating S.D. for change in walking speed *the 1/4 rule*

- Range of changes over 3 years*
- +0.2 m/sec to -0.6 m/sec
  - Based on 'experience' or informed guess
- Range = 0.8 m/sec
- 1/4 of the range = 0.2 m/sec

E/S

- Effect size: 0.1 m/sec difference in change
- Standard deviation: 0.2 m/sec
  - (We also know this from our cohort studies)
- E/S = 0.5

The New Ingredients

New Testable Hypothesis

Type of Study: Analytical (RCT)

Statistical Test: T-test

Effect Size: 1.0 vs 0.1 m/sec; E/S = 0.5

Power: 0.80; Alpha: 0.05

Sample Size Required per Group When Using the *t* Test to Compare Means of Continuous Variables

TABLE 6A Sample Size *per Group* for Comparing Two Means

One-sided α =	0.005			0.025			0.05		
Two-sided α =	0.01			0.05			0.10		
*E/S** β =	0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20
0.10	3,565	2,978	2,338	2,600	2,103	1,571	2,166	1,714	1,238
0.15	1,586	1,325	1,040	1,157	935	699	963	762	551
0.20	893	746	586	651	527	394	542	429	310
0.25	572	478	376	417	338	253	347	275	199
0.30	398	333	262	290	235	176	242	191	139
0.40	225	188	148	164	133	100	136	108	78
0.50	145	121	96	105	86	64	88	70	51
0.60	101	85	67	74	60	45	61	49	36
0.70	75	63	50	55	44	34	45	36	26
0.80	58	49	39	42	34	26	35	28	21
0.90	46	39	21	34	27	21	28	22	16
1.00	38	32	26	27	23	17	23	18	14

* *E/S* is the standardized effect size, computed as *E* (expected effect size) divided by *S* (SD of the outcome variable).

The New Ingredients

New Testable Hypothesis

Type of Study: Analytical (RCT)

Statistical Test: T-test

Effect Size: 1.0 vs 1.1 m/sec; E/S = 0.5

Power: 0.80; Alpha: 0.20

- **Sample size: 64 per group; 128 total**
- **With 20% drop out: 160 total**

Improving Precision

- Increased precision (decreased SD) will decrease the sample size
- Ways to increase precision:
  - Take the mean of repeated 6m walks
  - Standard deviation improves from 0.2 to 0.15 m/sec
- E/S improves from 0.5 to 0.1÷ 0.15 = 0.7

Sample Size Required per Group When Using the *t* Test to Compare Means of Continuous Variables

TABLE 6A Sample Size *per Group* for Comparing Two Means

One-sided α =	0.005			0.025			0.05		
Two-sided α =	0.01			0.05			0.10		
*E/S** β =	0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20
0.10	3,565	2,978	2,338	2,600	2,103	1,571	2,166	1,714	1,238
0.15	1,586	1,325	1,040	1,157	935	699	963	762	551
0.20	893	746	586	651	527	394	542	429	310
0.25	572	478	376	417	338	253	347	275	199
0.30	398	333	262	290	235	176	242	191	139
0.40	225	188	148	164	133	100	136	108	78
0.50	145	121	96	105	86	64	88	70	51
0.60	101	85	67	74	60	45	61	49	36
0.70	75	63	50	55	44	34	45	36	26
0.80	58	49	39	42	34	26	35	28	21
0.90	46	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	16
1.00	38	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	14

* *E/S* is the standardized effect siz … ble).

A modest improvement in precision reduced sample size by 1/2

Summary

- Estimate sample size early
- Systematically collect the ingredients
- Effect size is the most difficult - and important - judgment
- Alternatives that reduce sample size
  – Compromise power
  – Increase effect size
  – Precise continuous outcomes

Steven R. Cummings, MD

Dr. Cummings is Professor of Medicine and of Epidemiology and Biostatistics Emeritus, and Director of the San Francisco Coordinating Center, which specializes in coordinating multi-center studies and clinical trials. He is a co-author of the textbook "Designing Clinical Research", and an elected member of the Institute of Medicine of the National Academy of Sciences. A general internist, Dr. Cummings is an international expert in osteoporosis and healthy aging.

Profile

FACULTY
Designing Clinical Research

The Research Question
Thomas Newman, MD

Sample Size and Power
Steven Cummings, MD

Subjects and Variables
Douglas Bauer, MD

Statistics
Charles McCulloch, PhD

Types of Study Designs
Kirsten Bibbins-Domingo, MD, PhD

Qualitative Research and Questionnaire Design
Miriam Kuppermann, PhD, MPH

Causal Inference
Deborah Grady, MD, MPH

Ethical Issues in Clinical Research
Bernard Lo, MD

ACKNOWLEDGMENTS

Designing Clinical Research ONLINE

Course Director: Deborah Grady, MD, MPH

Course Designer: Chrisanne Garrett, MAED

Director, Online Education

chris.garrett@ucsf.edu

CTSI.ucsf.edu | Accelerate: Services for Researchers | UCSF Profiles

Follow CTSI
Twitter | SlideShare | YouTube | Flickr