

# Designing Clinical Research: An Introduction

Derek Stephens, Biostatistician  
Hospital For Sick Children  
Biostatistics: Design and analysis  
October 6, 2016

---

---

# Anatomy of research

What it's made of

---

# Anatomy of research

---

- ❑ Structure of a research project/clinical trial is set out in its protocol, the written plan of the study
- ❑ Helps the investigator organize their research in a logical, focused, and efficient way

# Success Criteria

---

- ❑ The protocol will go through many iterations, criticisms, and rejections.
  - ❑ Must be flexible while maintaining the primary study question(s).
  - ❑ Not the application of statistical tests, although the design will always involve an analysis plan.
-

# Anatomy of research

---

- Tangible elements of the study plan
  - Research question (Hypothesis)
  - Literature history
  - Design
  - Subjects(inclusion/exclusion criteria)
  - Measurements(outcomes and IV's)
  - Analysis and Sample size
  - Budget

# Anatomy of research

---

- Efficient

- ☐ Will the type of design be able to answer the question without needing a huge number of patients?
- ☐ Will the study be able to control major sources of variation?

# Anatomy of research

---

- Feasible

- ☐ Do the study have the appropriate measurement scales, personnel, and required subjects -number and heterogeneity- to be able to answer the hypothesis being tested?
- ☐ Is the time scale to do the study adequate?

# Anatomy of research

---

- Cost-effective
  - Will the cost incurred balance the gains from the study?
  - How many patients do you need to treat to prevent 1 “bad” outcome?
- [https://www.youtube.com/watch?feature=player\\_embedded&v=9Jpcw2E-Omw](https://www.youtube.com/watch?feature=player_embedded&v=9Jpcw2E-Omw)



# Outline of a study protocol

---

## □ Section

■ Research questions

■ Background and significance

■ Design

■ Analysis

## □ Purpose

■ What questions will the study address?

■ Why are these questions important?

■ minimize bias, reduce variance

■ Statistical (error in null rejection)

■ Estimates  $\pm$  SD

# 5 essential characteristics of a good research question

---

## □ Acronym FINER

- Feasible
- Interesting
- Novel
- Ethical
- Relevant

# Sections/Elements of the research protocol

---

- ❑ We will address these components in the context of a specific protocol.
- ❑ Effect of Liposomal Lidocaine and Sucrose Alone and in Combination for Venipuncture Pain in Newborns
- ❑ Anna Taddio, Vibhuti Shah, Derek Stephens et. al.
- ❑ Pediatrics 03/2011

# What is the problem to be addressed? Research Question

---

- ❑ Pharmacological interventions that **reduce procedural pain** have been prioritized in neonatal pharmacology research by international government agencies including the FDA.
  - ❑ Procedural pain in newborn infants is a significant burden to infants, their parents, healthcare workers, and society at large. Despite the fact that modern medical practice is greatly advanced, ***pain is inflicted on all Canadian newborn infants*** in the first days and months of life from blood tests, injections and cannulations designed to prevent, diagnose or manage medical conditions.
-

# Background and significance

---

- ❑ Sets the proposed study in context
- ❑ Gives its rationale
  - What is known about the topic at hand?
  - Why is the research question important?
- ❑ Cites previous research of relevance
- ❑ Indicates problems with prior research and what uncertainties remain

#### What is the problem to be addressed?(continued)

---

- ❑ Pain from these procedures is harmful. Acutely, pain leads to distress and physiologic instability.
  - ❑ Long-term, it alters neuronal circuitry and leads to increased future reactivity to pain, pre-procedural anxiety due to conditioning, needle phobia and avoidance of medical treatment.
  - ❑ It impacts on parents and healthcare workers also; both feel **stress** when infants and children are subjected to procedural pain and have concerns about the long-term adverse effects on children and themselves. Parents report infant suffering as the 'worst thing' about the hospital and maintain they would pay for analgesics.
-

#### What is the problem to be addressed?(continued)

---

- In summary, ***all young infants experience procedural pain***. Procedural pain, if left untreated, ***can be harmful*** to infant development. **There are currently no widely accepted analgesic regimens for the prevention of procedural pain.**
  - Liposomal lidocaine and sucrose are good candidates for widespread clinical use but additional pharmacological evaluation of these agents is required.
  - This study will test the ***relative and combined efficacy of liposomal lidocaine and sucrose and the safety of liposomal lidocaine***. The results will be used to ***develop and disseminate protocols for procedural pain management***.
-

# Design

---

- ❑ The design should describe what was previously done in regards to this question
- ❑ The design should lay out how the question(s) will be answered.
- ❑ Should address potential biases that could lead to systematic errors.
- ❑ Should describe methods to overcome these biases



# Major Sources of Bias in Research Studies

---

- There are two types of error associated with most forms of research:
    - Random and Systematic.
    - Both random and systematic errors can threaten the validity of any research study.
-

# Measurement Error

---

$$X = T + e$$

*Two Components:*

$e_r$  • Random Error

$e_s$  • Systematic Error

$$X = T + e_r + e_s$$

# Error

---

## Random errors.

Due to sampling variability or measurement precision.

Adds variability to the data, but does not affect the average performance of the group.

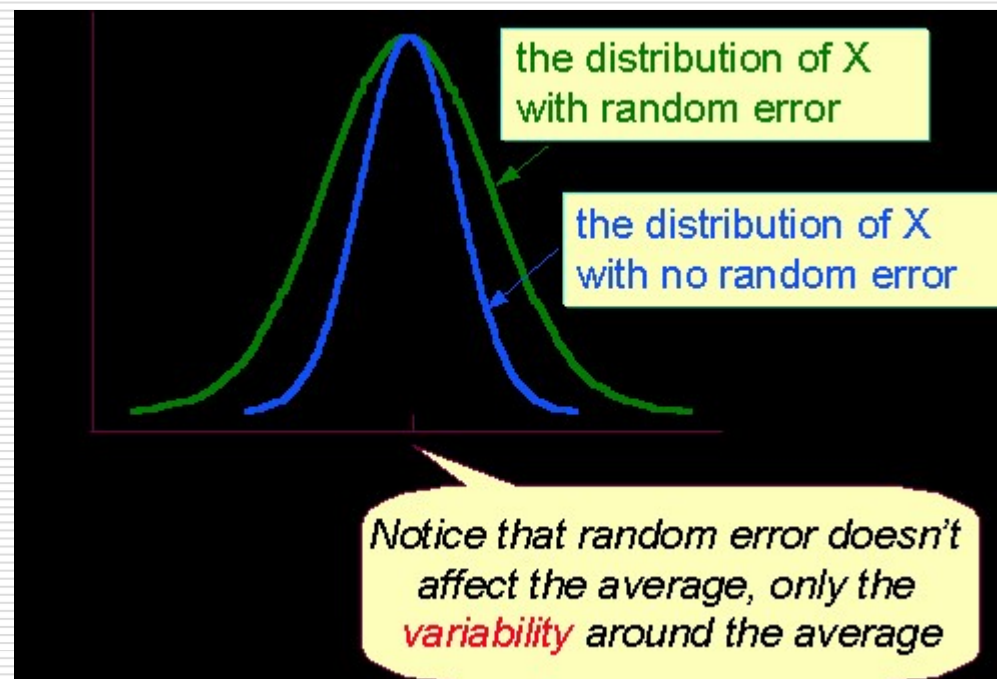
Occurs in essentially all quantitative studies and can be minimized but not avoided.

(minimized by increasing sample size, inclusion of SOP's, random assignment of patients to groups, etc.)

---

# Random error

---



# Error

---

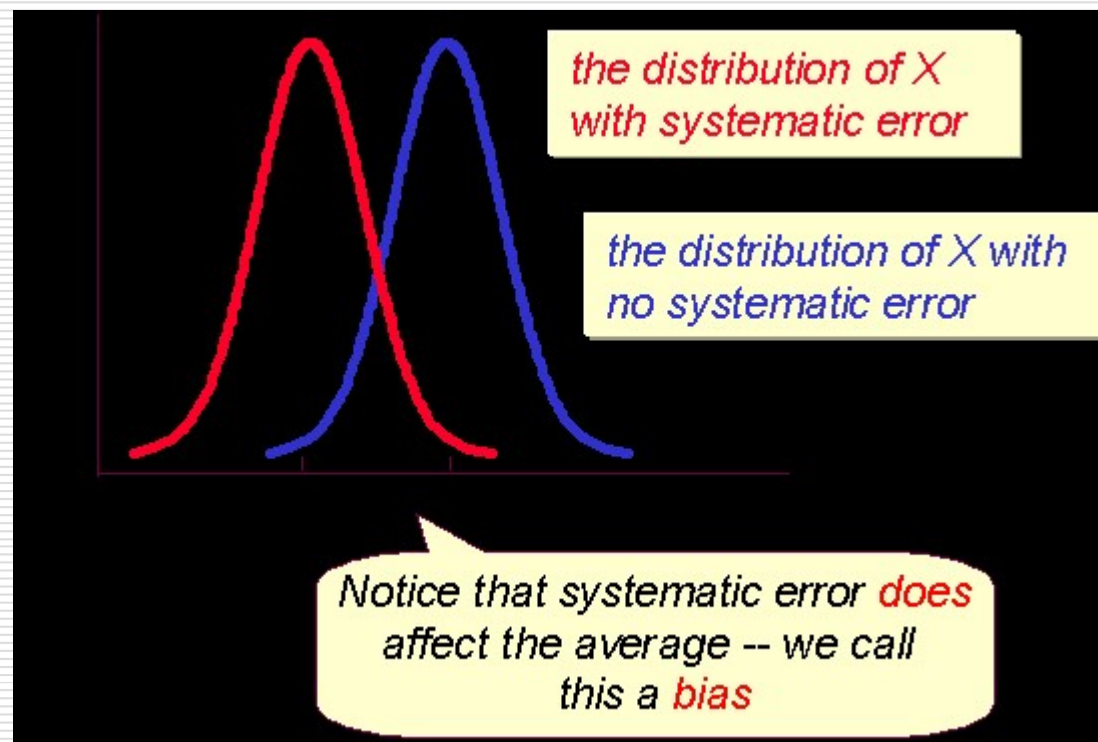
## Systematic errors.

inaccuracies that produce a consistently false pattern of differences between observed and true values.

Systematic error is caused by any factor(s) that systematically affect measurement of the variable across the sample. Unlike random error, systematic errors tend to be consistently either positive or negative -- because of this, systematic error is sometimes considered to be ***bias*** in measurement.

# Systematic error

---



# Bias

---

- ❑ Random errors can be determined and addressed using statistical analysis.
  - ❑ Most systematic errors/biases cannot. This is because biases can arise from innumerable sources, including complex human factors.
  - ❑ Avoidance of systematic errors/biases is the task of a good research design
-

# Major Categories of Research Bias

---

- There are many different types of biases described in the research literature. The most common categories of bias that can affect the validity of research include the following:
    1. *Selection biases*, which may result in the subjects in the sample being unrepresentative of the population of interest
    2. *Measurement biases*, which include issues related to how the outcome of interest was measured
    3. *Intervention (exposure) biases*, which involve differences in how the treatment or intervention was carried out, or how subjects were exposed to the factor of interest
-



## Selection Biases

---

### □ Selection biases.

- Occurs when the groups to be compared are different. These differences may influence the outcome. Common types of sample (subject selection) biases include volunteer or referral bias, and nonresponder bias.
  - So missing data could become a problem, unless the missing data is at random.
-

## Selection Biases

---

### ☐ *Volunteer or referral bias.*

occurs because people who volunteer to participate in a study (or who are referred to it) are often different than non-volunteers/non-referrals. This bias usually, but not always, favors the treatment group, as volunteers tend to be more motivated and concerned about their health.

---

# Selection Biases

---

## □ *Nonrespondent bias.*

- those who do not respond to a survey differ in important ways from those who respond or participate. This bias can work in either direction.
-

# Measurement Biases

---

- ❑ Measurement biases involve systematic error that can occur in collecting relevant data. Common measurement biases include instrument bias, insensitive measure bias, expectation bias, recall or memory bias, attention bias, and verification or work-up bias.
-

# Measurement Biases

---

## *Instrument bias.*

when calibration errors lead to inaccurate measurements being recorded, e.g., an unbalanced weight scale.

## *Insensitive measure bias*

when the measurement tool(s) used are not sensitive enough to detect what might be important differences in the variable of interest.

## *Expectation bias.*

Expectation bias occurs in the absence of masking or blinding, when observers may err in measuring data toward the expected outcome. This bias usually favors the treatment group

---

# Measurement Biases

---

## □ *Recall or memory bias.*

- subjects recall past events.
- Often a person recalls positive events more than negative ones.
- Alternatively, certain subjects may be questioned more vigorously than others, thereby improving their recollections.

## □ *Attention bias.*

- people who are part of a study are usually aware of their involvement, and as a result of the attention received may give more favorable responses or perform better than people who are unaware of the study's intent.
-

# Intervention (Exposure) Biases

---

## □ *Contamination bias.*

- members of the 'control' group inadvertently receive the treatment or are exposed to the intervention, thus potentially minimizing the difference in outcomes between the two groups.

## □ *Co-intervention bias.*

- subjects are receiving other (unaccounted for) interventions at the same time as the study treatment.
-

# Intervention (Exposure) Biases

---

## □ *Timing bias(es).*

- If an intervention is provided over a long period of time, maturation alone could be the cause for improvement. If treatment is very short in duration, there may not have been sufficient time for a noticeable effect in the outcomes of interest.

## □ *Compliance bias.*

- when differences in subject adherence to the planned treatment regimen or intervention affect the study outcomes..



# Design Considerations-What features to adopt.

---

- ❑ Decision made by the PI:
  1. Take a passive role in observing events in study subjects
    - Observational study(follow patients who come to clinic)
  2. Apply an intervention and examine its effects
    - Clinical trial

# The Proposed Design

---

**What is the proposed trial design?**

- ☐ This is a ***double-blind, randomized, controlled, double-dummy, single-centre trial*** involving 330 full-term newborn infants in the well baby nursery undergoing the newborn screening test.
- ☐ Is the trial/sample size feasible?

## Components of the design

### What are the planned trial interventions (experimental and control)

---

- ❑ Baseline characteristics will be collected and infants will be *randomized* (*is this ethical*) to one of X possible regimens administered by the nurse assigned to the care of the infant: (1) 1g of liposomal lidocaine to the dorsum of the hand for 30-40 minutes prior to venipuncture, occluded by a dressing (Tegaderm™); (2) 2ml of 24% sucrose (administered by mouth using a syringe over 1-2 minutes), 2 minutes prior to venipuncture; (3) both liposomal lidocaine and sucrose.
  - ❑ Venipunctures will be performed by a variety of *trained personnel* (to minimize intervention bias), as per usual clinical practices (i.e., registered nurses or neonatology trained physicians) in the treatment room on the postnatal floor, the usual setting, thus standardizing the process of venipunctures.
  - ❑ We *previously* demonstrated efficacy of these doses and administration techniques. Identical appearing placebos will be used for liposomal lidocaine and sucrose (i.e., double-dummy), so that all infants receive liposomal lidocaine or placebo and all infants receive oral sucrose or placebo (water).
  - ❑ *What is double-dummy?*
-

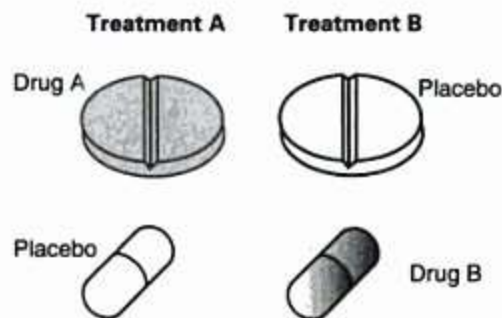
# Double-Dummy

---

- ❑ Double dummy is a technique for retaining the blind when administering treatments in a clinical trial, when the two treatments cannot be made identical. Topical cream and liquid sucrose cannot be made identical.
  - ❑ Supplies are prepared for Treatment A (active and indistinguishable placebo) and for Treatment B (active and indistinguishable placebo). Subjects then take one of two sets of treatment; {either A (active) and B (placebo)} or {A (placebo) and B (active)}.
-

# Double-Dummy technique

---



**Fig. 2.7** The double-dummy technique. The patient always takes a tablet and a capsule. In treatment A, the tablet contains the active drug and the capsule contains the placebo. In treatment B, the capsule contains the active drug and the tablet contains the placebo.

# What is a double-blind study?

---

- In a randomized double-blind, placebo-controlled trial of a medical treatment, some of the participants are given the treatment, others are given fake treatment (placebo), and **neither the researchers nor the participants** know which is which until the study ends (they are thus both “blind”). The assignment of participants to treatment or placebo is done randomly, perhaps by some random generating mechanism.
-

## What are the proposed methods for protecting against sources of bias?

---

- ❑ A double-blind, double-dummy, randomized, controlled design will minimize bias. (Expectation bias)
  - ❑ Blinding to group assignment will be ensured since the containers with the interventions are made visually indistinguishable from each other.
  - ❑ Bias during data collection will be minimized by ensuring that study personnel and clinical staff are unaware of group assignment (Expectation bias).
  - ❑ Outcomes will be assessed by trained study personnel ( instrument bias) unaware of group assignment (Expectation bias) or study hypotheses.
-

# Common observational study designs(more on this next class)

---

## ☐ Cohort

- Prospective(assemble now follow in time)
- Retrospective (look back in time)

## ☐ Cross-sectional (at a single time)

## ☐ Case-control (Cases/controls defined first)

*Was the Lidocaine pain study just described an observational study?*



# Choice of design

---

- ❑ No single approach is always(i.e. under any circumstance) better than the others.
- ❑ The randomized double blind trial is often cited as best for establishing causality, but it's not always *feasible*

# Chronology of study designs

---

## ☐ Descriptive

### ■ Explore the lay of the land

☐ E.g. Estimate proportions of subjects experiencing pain and health-related characteristics in the population.

■ What is the average number painful venipunctures in different subgroups of infants i.e. diabetic / healthy.

# Chronology of study design

---

## ☐ Analytic

- Evaluate associations to permit inferences about cause-and-effect relationships
  - ☐ E.g. Is there an association between treatment and pain?

*Pain study will address both issues descriptive and analytic.*

# Chronology of study design

---

## ☐ Clinical trial

- Occur relatively late in a series of research studies because they tend to be more difficult and expensive
- Tend to answer questions more definitively
- Answer questions more narrowly focused that arise from the findings of observational studies

## ☐ The current protocol had all in one

---

# Clinical trial(Summary)

---

- Principal Question: In full-term newborn infants undergoing a venipuncture for the newborn screening test: (1) What is the analgesic efficacy of sucrose alone, liposomal lidocaine alone, and sucrose plus liposomal lidocaine as assessed by behavioural and physiologic responses? (*upto now no outcome variables have been discussed*)
- Secondary Questions: (2) Does liposomal lidocaine result in toxicologically significant plasma lidocaine levels [defined as >1mcg/ml, which is 20% of the value associated with clinical toxicity (i.e. 5mcg/ml considered toxic)] (to ensure patient safety )
- **We hypothesize** that: (1) sucrose plus liposomal lidocaine will be superior to either agent alone in reducing pain during venipuncture; (2) plasma levels of lidocaine will be below toxicologically significant levels (1 mcg/ml)

# Study subjects

---

- ❑ Inclusion and exclusion criteria that define the target population
  - The *kinds* of patients best suited to the research question (full-term newborn infants undergoing a venipuncture for the newborn screening test) ( to minimize selection bias)
- ❑ How to recruit the study sample
  - E.g. identify subjects in a clinic with relevant diagnosis codes.

**Which variables should be *measured*?**  
**What are the proposed primary and secondary outcome measures?**

---

- The **primary outcome measure** is infant pain during venipuncture for the newborn screening test, as assessed by facial grimacing response.
- Facial grimacing is a *validated* (to control insensitive measure bias) measure of acute pain in neonates and considered the most sensitive and specific marker of pain in newborn infants.

## Premature Infant Pain Profile (PIPP)

---

- ☐ gestational age
- ☐ behavioral state before painful stimulus
- ☐ change in heart rate during painful stimulus
- ☐ change in oxygen saturation during painful stimulus
- ☐ brow bulge during painful stimulus
- ☐ eye squeeze during painful stimulus
- ☐ nasolabial furrow during painful stimulus

Stevens B Johnston C et al. Premature Infant Pain Profile: Development and initial validation. Clinical Journal of Pain. 1996; 12: 13-22



## **Which other variables should/should not be measured?**

### **What are the proposed primary and secondary outcome measures?**

---

- The Premature Infant Pain Profile (PIPP), a composite measure of pain which incorporates 7 components of facial grimacing response (brow bulge, eyes squeezed shut, naso-labial furrow)
  - physiological response (heart rate, oxygen saturation) and contextual variables (gestational age, state) was not selected because divergence between behavioural (facial grimacing) and physiologic responses (heart rate) have been previously reported. Moreover, in our previously-funded trial (MCT-63143), we observed an increase in heart rate following administration of sucrose that persisted throughout the painful procedure, questioning the validity of heart rate as a measure of analgesic effectiveness of sucrose.
  - Facial grimacing forms the basis of most composite measures, including the PIPP, and is the best validated unidimensional measure of pain.
  - Has been shown to be a sensitive measure of infant pain.
-

## Variables (specification and validation)

---

- The facial grimacing score will incorporate three facial actions (brow bulge, eye squeezed shut, naso-labial furrow) that are individually recorded as present or absent during a specified time interval (described in section 2.9 of the protocol). We have utilized this method of pain assessment in our previous studies.
-

## Variables(continued)

---

- **Secondary outcomes** include visual analog scale scores, cry duration, heart rate, number of attempts until procedure completion, procedure duration.
  - note some of these may be *confounding* variables.
-

# Variables (in summary)

---

- Validated and Reliable Outcomes
- Other factors that may influence pain responses: infant strata (with or without IV cannula), mode of delivery, gender, race, previous painful procedures, blood pressure, feeding pattern, maternal antidepressant use
- Intervention i.e. groups
- Confounders

## What is the proposed *sample size* and what is the justification for the assumptions underlying the power calculations?

---

- The sample size is based on the ability to detect a **clinically important difference** in efficacy among analgesic regimens, whereby a change of  $\geq 0.18$  points [standard deviation (SD)=0.35] in the facial grimacing score, or  $\sim 0.5SD$ , is considered a minimally clinically important difference. Hence the plan is to test all groups to each other so there will be 3 post-hoc comparisons. In consideration of multiple testing, the level of significance will be 1.7% (i.e., **Type I error rate,  $\alpha=0.017$** ). **Power will be set at 90%**. The smallest estimated difference (i.e., 0.18) is between sucrose and sucrose plus liposomal lidocaine. Other differences are expected to be larger with a similar standard deviation, hence sample size is based on the smallest difference. Larger effect sizes require fewer subjects so power for the 2 other differences will be  $>90\%$ . The estimated mean pain scores for the sucrose and sucrose plus liposomal lidocaine groups (obtained from unpublished and published data) are: for sucrose = 0.22 (SD=0.35); sucrose plus lidocaine = 0.04 (SD=0.35). In consideration of these estimates, a sample size of 103/group is required. Hence the total sample size is  $(3 \times 103) = 309$ . [i] To account for dropouts, 110 infants/group (total=330) will be enrolled.
  - Power and samples size (PASS) Computer Software. NCSS; Utah: 2000.
-

# Statistical Analysis

---

- ❑ **Principal Question**: (1) The relative efficacy of liposomal lidocaine, sucrose, and liposomal lidocaine plus sucrose will be analyzed using a one way ANOVA that compares pain scores among groups during the needle puncture phase of the venipuncture. Subsequently, pair wise comparisons between groups will be performed; the p-value will be adjusted to account for multiplicity of tests ( $\leq 0.017$  will be considered significant).
- ❑ GLM will be used to assess confounding/adjustment for other factors (if needed)
- ❑ The secondary analysis will use repeated measures using MIXED analysis that compares pain profiles among the groups for the entire procedure.

# Statistical Analysis

---

- Baseline characteristics and secondary outcomes will be analyzed using t-tests, one-way ANOVA or  $\chi^2$ -tests, according to the type of data (continuous or categorical). Non-parametric tests will be considered when appropriate. An *intention-to-treat analysis* as well as a *protocol compliant* analysis will be performed. The statistical package to be used is SAS (V.9, Cary, NC).

# 'Intention-to-treat Analysis'

---

- method of analyzing results of a randomized controlled trial that includes in the analysis all those cases that should have received a treatment regimen but for whatever reason did not. All cases allocated to each arm of the trial are analyzed together as representing that treatment arm, regardless of whether they received or completed the prescribed regimen
-



# (adjustment for) Confounding variables

---

- ☐ *In the simplest experiment, one investigates the relationship between two things by deliberately producing change in one of them and...observing the change in the other.*
- ☐ *Frequently "something else" gets included and masks the true causal relationship.*

## Confounding variables (continued)

---

- The “something else” would be a confounding variable, defined as “an unforeseen and unaccounted-for variable that jeopardizes the reliability and validity of an experiment's outcome.”

# Possible confounding variables

---

- Babies crying prior to treatment
  - Previous painful procedures
  - Blood pressure
  - Feeding pattern
  - Maternal antidepressant use
- 
- Would be more of a concern in a non-randomized trial

## Confounding variables (continued)

---

- ☐ if subjects were taken from different clinics say the diabetes clinic or the asthma clinic, this could be a confounding variable? (or introduce a co-intervention bias)
- ☐ Patients with diabetes may be more conditioned to getting insulin injections, and so this conditioning may produce different results.
- ☐ So Clinic may be a potential confounder

# Systematic error

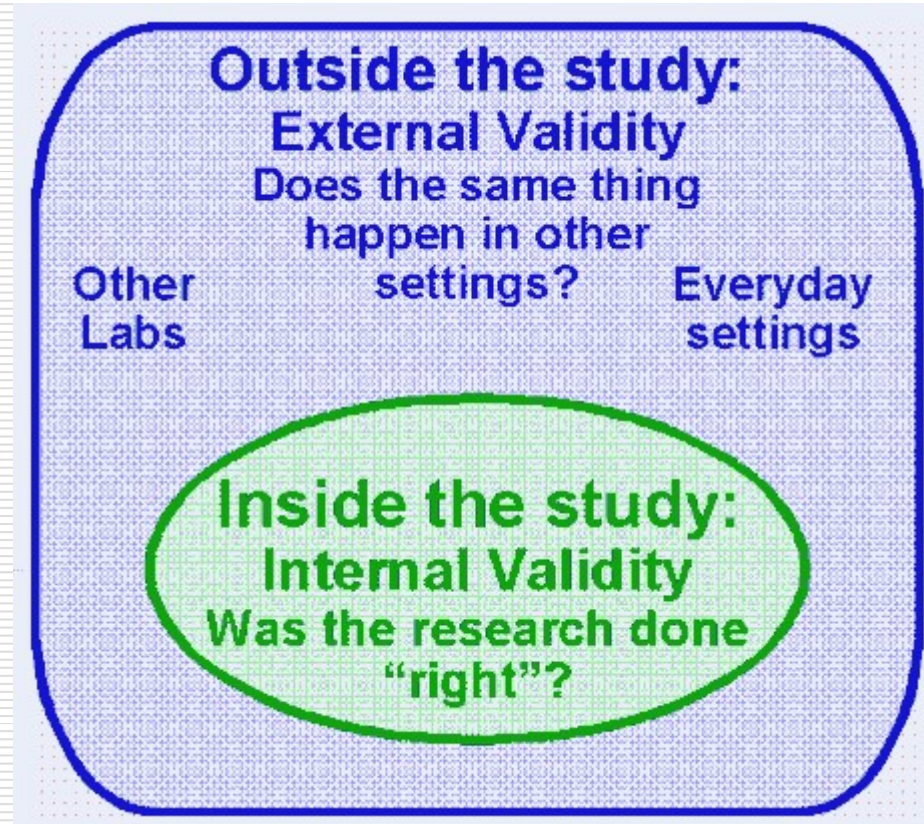
---

- ❑ Misleading/incorrect result due to one of the many biases discussed.
- ❑ Increasing the sample size has no effect on reducing systematic error.

*(To control systematic errors need to)  
Improve the design by minimizing the  
biases involved.*

# Internal/External Validity

---



# Internal Validity

---

- ❑ ***Internal validity*** refers to how well an experiment is done, especially whether it avoids confounding and bias (more than one possible independent variable [cause] acting at the same time).
- ❑ The less chance for confounding and more control of bias in a study, the higher the internal validity.

# External validity

---

- ❑ The degree to which the conclusions can be appropriately applied to people and events outside the study
- ❑ This is also called generalizability
- ❑ Generally this is harder to establish since the study population may not represent the population of interest or the measurement of variables may not be as was intended, or there may be a lot of missing data etc.