BME 2200: BME Biostatistics and Research Methods

Lecture 3: Experimental design

Lucas C. Parra
Biomedical Engineering Department
City College of New York

CCNY

parra@ccny.cuny.edu

Content, Schedule

1. Scientific literature:
- Literature search
- Structure biomedical papers, engineering papers, technical reports
- Experimental design, correlation, causality.

2. Presentation skills:
- Report – Written report on literature search (individual)
- Talk – Oral presentation on biomedical implant (individual and group)

3. Graphical representation of data:
- Introduction to MATLAB
- Plot formats: line, scatter, polar, surface, contour, bar-graph, error bars. etc.
- Labeling: title, label, grid, legend, etc.
- Statistics: histogram, percentile, mean, variance, standard error, box plot

4. Biostatistics:
- Basics of probability
- t-Test, ANOVA
- Linear regression, cross-validation
- Error analysis
- Test power, sensitivity, specificity, ROC analysis

Correlation versus Causality

Consider hair length and testosterone levels. One may observe in a given sample of the population a strong correlation (in our culture testosterone correlates strongly with short hair). But this does imply a mechanistic or causal link between the two:

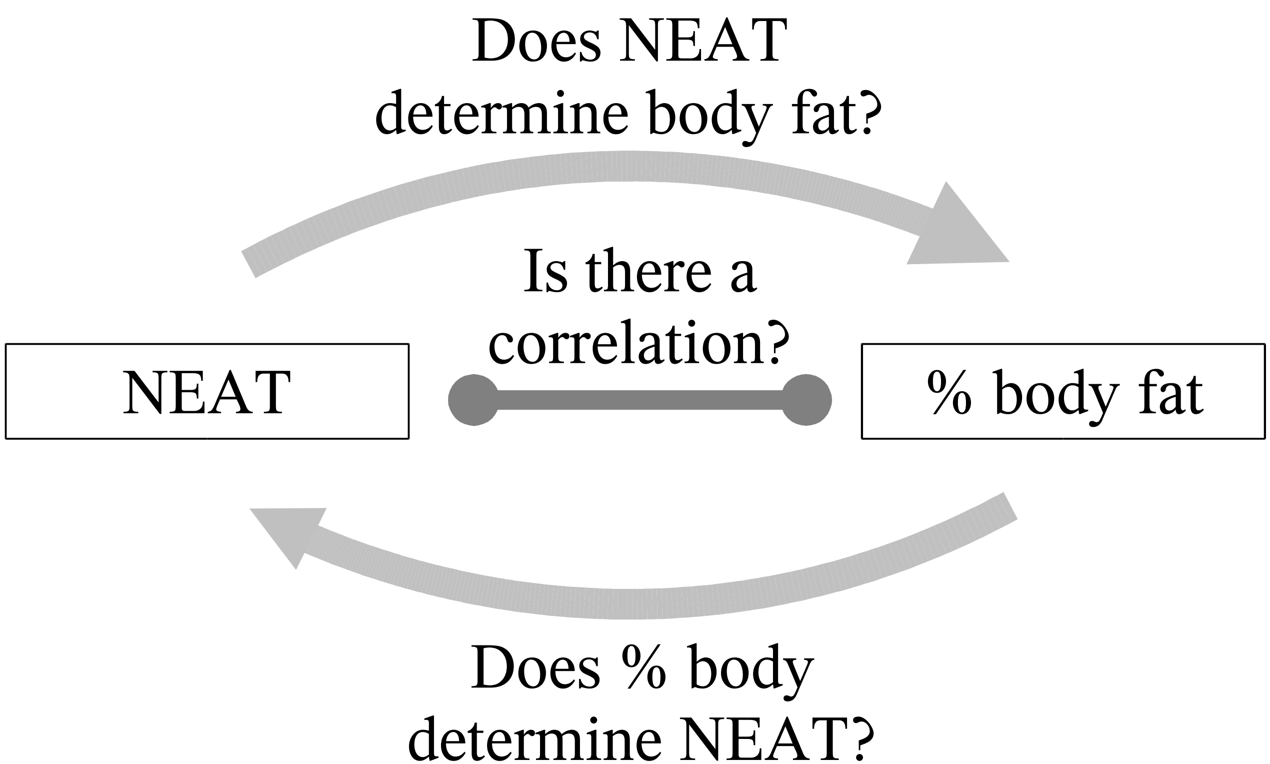
Cutting ones hair will not increase testosterone levels, nor will a testosterone injection reduce the length of your hair (if anything it will cause the opposite).

Correlation does not imply causality!

To test causality on has to be able to **manipulate** one of the variables (the **independent variable**) and **observe** how this affects the other variable (**dependent variable**).

Correlation versus Causality

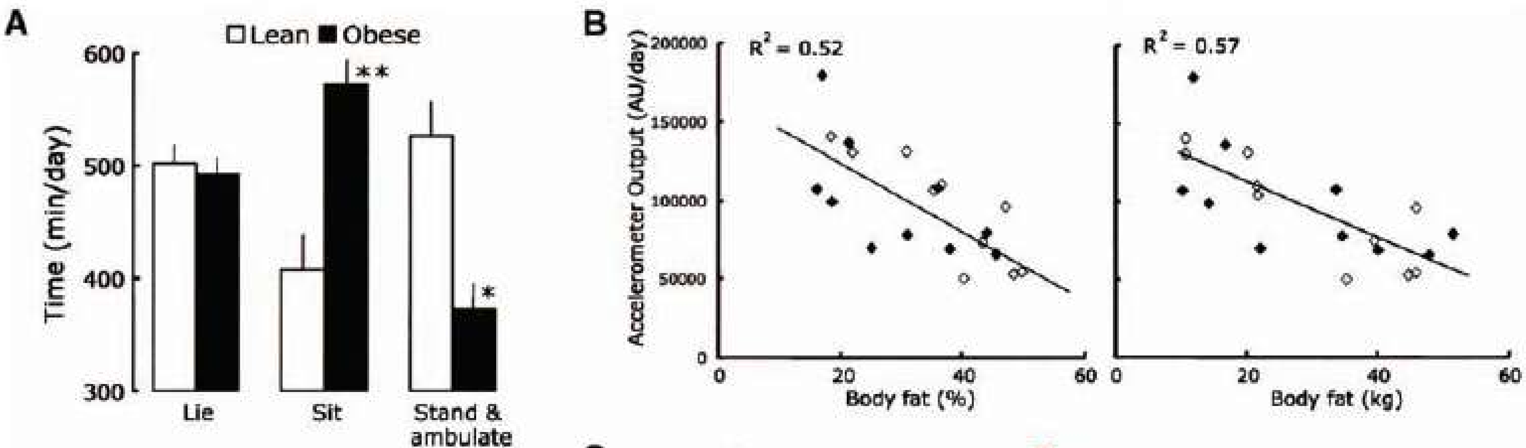
Which of these questions on NEAT behavior and percent body fat does the paper by Levine et al. *Science* 2005 address?



Which were Levine's dependent and independent variables?
What experiment would answer the missing link?
What would be the corresponding independent/dependent variable?

Correlation versus Causality

Extracted from Levine et al., *Science,* 307 (5709): 584-86, 2005



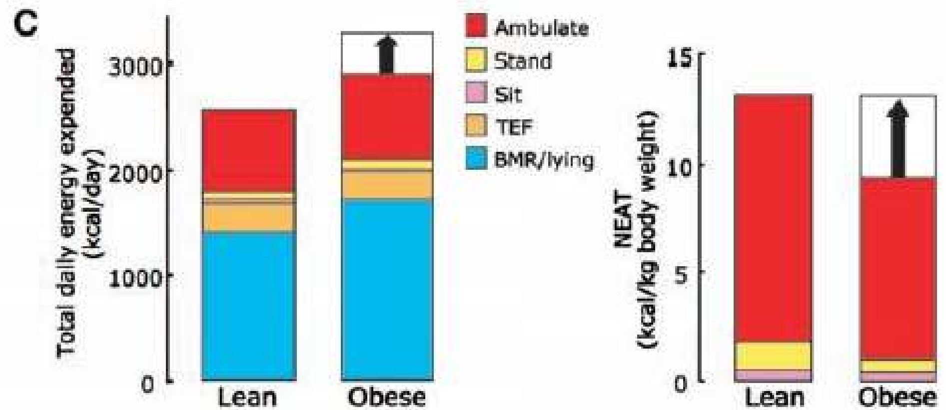


Fig. 1. (**A**) Time allocation for different postures for 10 obese and 10 lean sedentary subjects. Data are shown as mean + SEM. Significant differences between lean and obese are indicated: *, *P* = 0.001; **, *P* = 0.0005. There were no statistically significant differences between females (*n* = 10) and males (*n* = 10): Females stood 470 ± 35 min/day and males stood 429 ± 40 min/day. (**B**) Relationship between total body movement and body fat content. Body fat, determined from dual x-ray absorptiometry, is expressed as a percentage (left) and mass (right) plotted against the total 10-day accelerometer output [accelerometer units (AU)] for 20 (10 obese and 10 lean) sedentary subjects. The open diamonds are data for females and the filled diamonds are data for males. There was no significant relationship between fat-free mass and accelerometer output (fig. S1). The relationship between NEAT by doubly labeled water adjusted by weight versus accelerometer output is shown in fig. S2. (**C**) (Left) Total daily energy expenditure and (right) NEAT in 10 obese and 10 lean sedentary subjects. The uppermost segments of the bars for obese individuals (vertical arrows) represent the additional energy that could be expended if these subjects were ambulatory for the same amount of time as lean subjects. BMR, basal metabolic rate; TEF, thermic effect of food. There was no significant difference in sleeping time between the lean group (423 ± 15 min) and the obese group (434 ± 17 min). The energy expenditure data and standard deviations appear in table S2. The relationship between NEAT measured with doubly labeled water and NEAT measured with the instruments is shown in fig. S3.

Correlation versus Causality

Extracted from Levine et al., *Science,* 307 (5709): 584-86, 2005

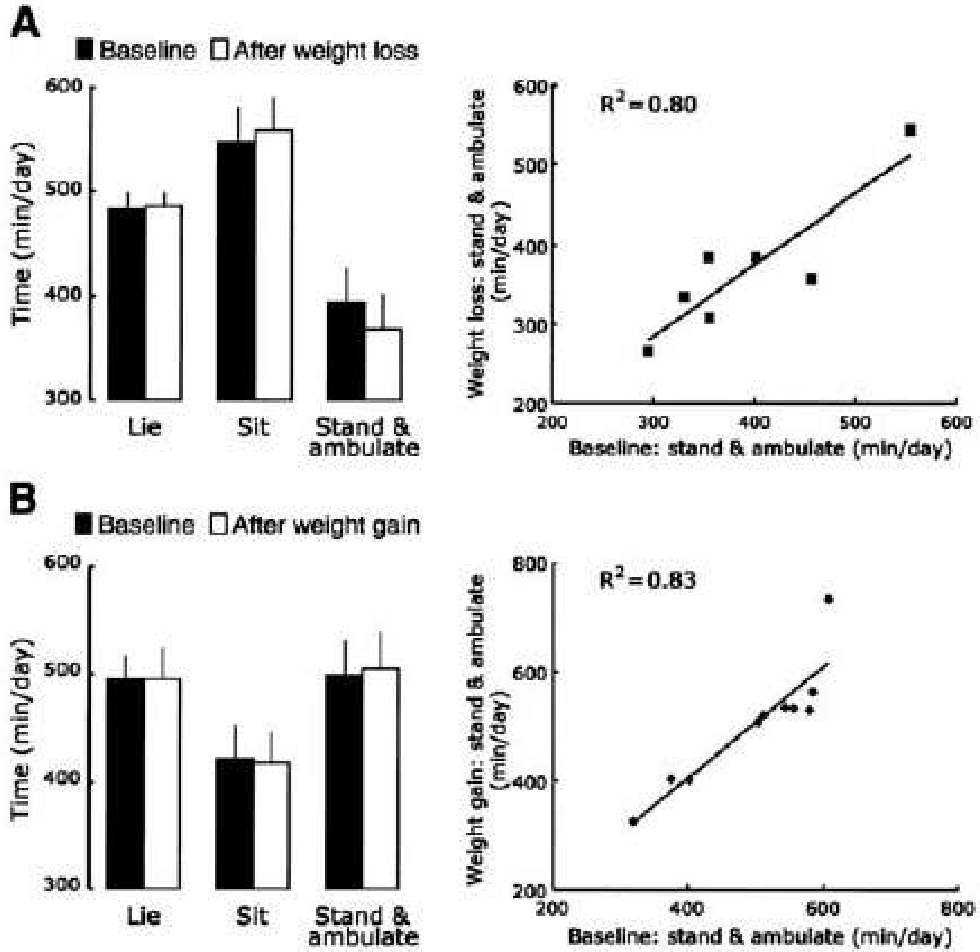
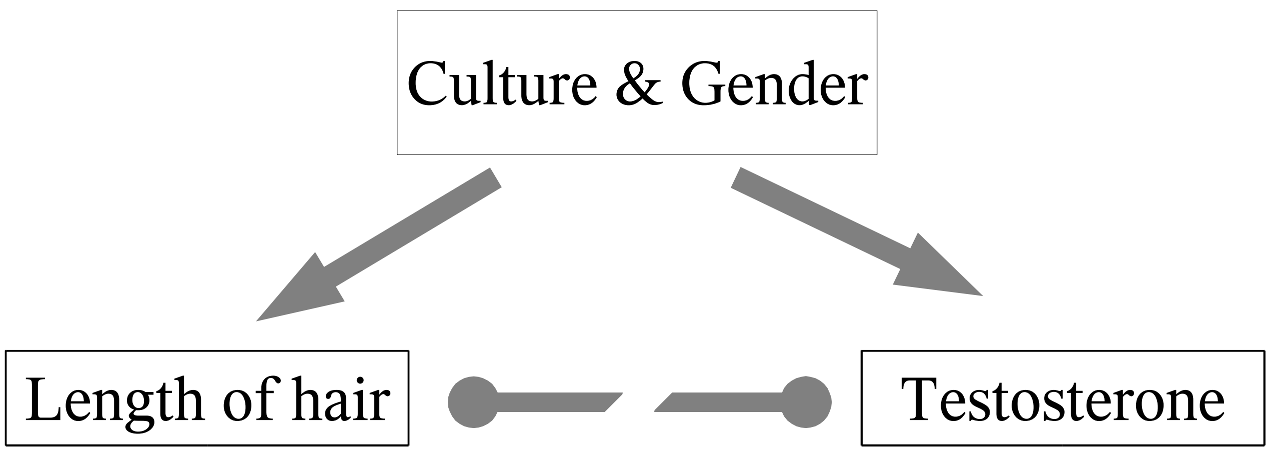


Fig. 2. (**A**) Posture allocation in seven obese sedentary subjects who underwent caloric restriction (*8*). (Left) Posture allocation data at baseline and after weight loss of 8 ± 2 kg. (Right) The time the subjects spent standing/ambulating at baseline is plotted against the time the subjects spent standing/ambulating after weight loss. (**B**) Posture allocation in 10 lean sedentary subjects who underwent experimental weight gain (*8*). (Left) The posture allocation data for baseline and after weight gain of 4 ± 2 kg. (Right) The time the subjects spent standing/ambulating at baseline is plotted against the time the subjects spent standing/ambulating after weight gain. Data are shown as mean + SEM.

Correlation without causality Causality?

When two variables are correlated, yet they do not causally affect one another it means there are one or more 'hidden' causes or unobserved conditioning variables, e.g.



Once we factor out the hidden variables we may find that there is no correlation. For instance by measuring correlation within a given gender, i.e. we **control** for gender **by fixing this variable**.

We call this '**conditional independence**'.

Experimental design

Observational study: It can only show association between variables but not causal dependency. (e.g. NEAT correlates with % body fat.)

Prospective Study: Can show causality (or disprove it). The independent variable is manipulated and one observes how the dependent variable changes as a result. (e.g. increasing body fat does not reduce NEAT)

Case-controlled study: Compares the dependent variable for two groups that are identical except for the independent variable. Can be done retrospectively. May show causality provided that the two groups are really "identical", i.e. Are all possible relevant variables matched/controlled?

Experimental design

Assignment 3:

1. Find a publication for one observational study, one prospective study, and one case-controlled study – ideally on the topic that you have selected for assignment 4.

2. Describe the relationship that is shown. Is the relationship only a correlation, or is it also a causality, and if so, which direction of causality?

3. Name the dependent variables, independent variables, and control variables.