

Principles of Experimental Design

Statistical Thinking in Biomedical Research

October 2, 2000

Mark Conaway

Use examples to illustrate principles

- Reference:

Maughan et al. (1996) Effects of Ingested Fluids on Exercise Capacity and on Cardiovascular and Metabolic Responses to Prolonged Exercise in Man. *Experimental Physiology*, **81**, 847-859.

- From paper summary

- “The present study examined the effects of ingestion of water and two dilute glucose-electrolyte drinks on exercise performance and
.....”

Process of Experimental Design

- What's the research question?
 - effect on exercise capacity...
- What treatments to study?
 - control group (no liquid intake) vs water vs 2 types of dilute glucose-electrolyte solutions
- What are the levels of the treatments?
 - Paper describes exact composition of solutions
- How to measure the outcome of interest
 - Exercise capacity: time to exhaustion on stationary cycle

Entire Process of Experimental Design

- Process of design relies heavily on researchers' knowledge of the field, though statistical principles can help
 - Do we need a “no liquid” control group?
 - Is “time to exhaustion” a valid measure of exercise capacity?

Statistical DOE: Allocate treatments to experimental material to...

- Remove systematic biases in the evaluation of the effects of the treatments
 - “unbiased estimates of treatment effects”
- Provide as much information as possible about the treatments from an experiment of this size
 - “precision”

Statistical DOE

- Remove bias, obtain maximum precision, keeping in mind
 - simplicity/feasibility of design
 - natural variation in experimental units
 - generalizability

Focus on comparative experiments

- Treatments can be allocated to the experimental units by the experimenter
- Other types of studies also have these as goals but:
 - Methods for achieving goals (unbiased estimates, precision) in comparative experiments rely on having treatments under control of experimenter

Back to example

- 4 treatments
 - no water (N)
 - water (W)
 - isotonic glucose-electrolyte(I)
 - hypotonic glucose-electrolyte (H)
- Outcome: time to exhaustion on bike
- Pool of subjects available for study

Design 1: subjects select treatment

- Does this method of allocation achieve the goals?
- Possible that this method induces biases in comparisons of treatments
 - e.g. Would “naturally” better athletes choose electrolytes?
 - e.g. Would more competitive athletes choose electrolytes?

Design 1A: Investigators assign treatments

- “Systematically”
 - Everyone on Monday gets assigned no water
 - Tuesday subjects get water only...
- “Nonsystematically”:
 - Whatever I grab out of the cooler...
- Again possible that this method induces biases in comparisons of treatments

What are the sources of the biases?

- Key point: Bias in evaluating treatments due to allocating different treatments to different types of subjects
 - e.g., “better” riders get electrolyte
 - so differences between treatments mixed up with differences between riders
- To have unbiased estimates of effects of treatment, need to have “comparable groups”

Randomization is key to having comparable groups

- Assign treatments at random
 - Note: Draw distinction between “random” and “non-systematic”
- Randomization is key element for removing bias
- In principle, creates comparable groups even on factors not considered by the investigator

Completely randomized design

- Randomly assign treatments to subjects
 - Generally assign treatments to equal numbers of subjects
- Does this give us the most information (precision) about the treatments?
- *Get precise estimates by comparing treatments on units that are as similar as possible.*

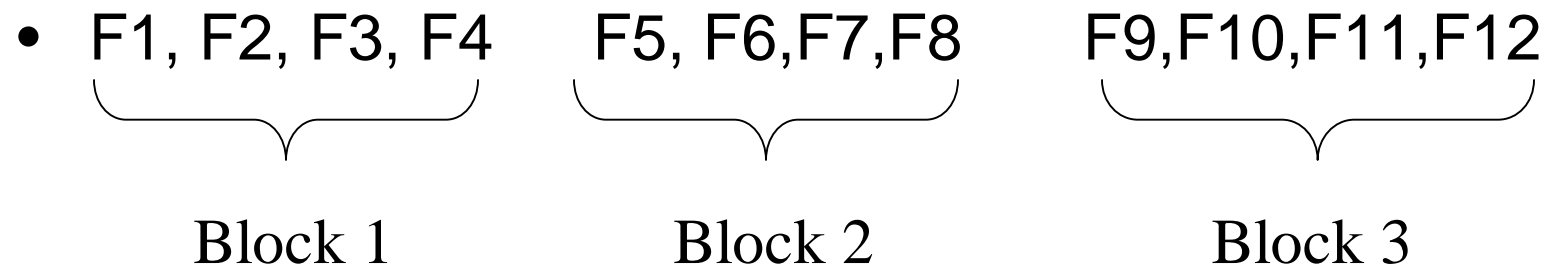
Randomized block designs (RBD)

General

- Group units into subgroups (blocks) such that units within blocks are more homogeneous than in the group as a whole
- Randomly assign treatments to units within subgroups (blocks)

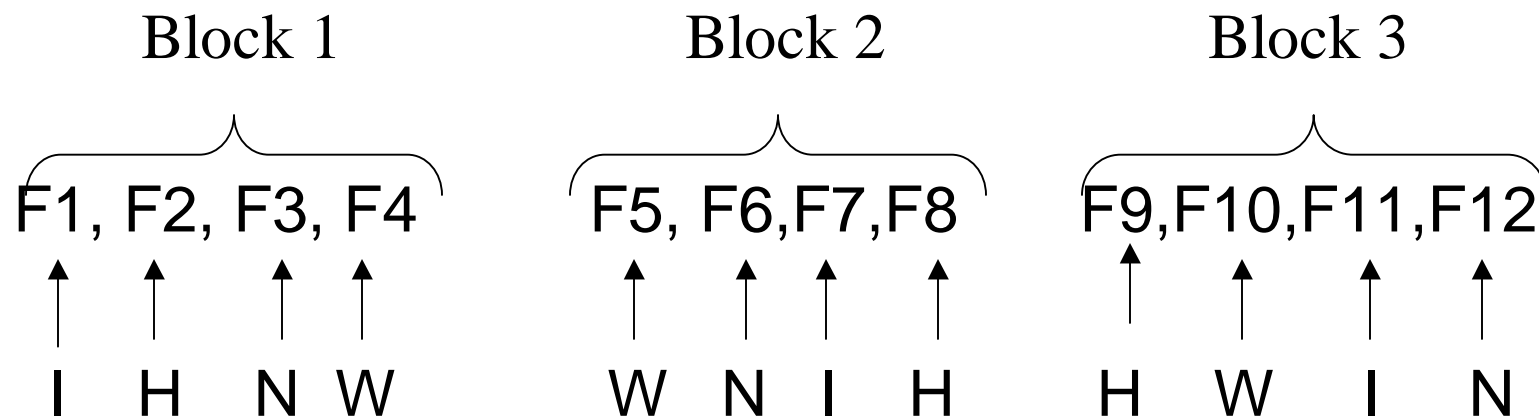
Randomized block designs in exercise example

- Do an initial “fitness screen” - let subjects ride bike (with water?) until exhaustion.
- Arrange subjects in order of increasing times (fitness)



Randomized block designs in exercise example

- Randomly assign treatments to units within



Advantages of RBD

- If variable used to create blocks is highly related to outcome, generally get much more precision than a CRD without doing a larger experiment
- Essentially guarantees that treatments will be compared on groups of subjects that are comparable on initial level of fitness

Disadvantages of RBD

- Now require 2 assessments per subject if block in this way
- Note: Could use some other measure of initial fitness that doesn't require an initial assessment on the bike

Can take idea further

- Could group by more than one variable
- Each blocking variable
 - Adds complexity
 - Might not increase precision if grouping variable is not sufficiently related to outcome

Repeated measures designs/Cross-over trials

- Natural extension of idea in RBD: want to compare treatments on units that are as similar as possible
- Subjects receive every treatment
- Most common is ``two-period, two-treatment''
 - Subjects are randomly assigned to receive either
 - A in period 1, B in period 2 or
 - B in period 1, A in period 2

Repeated measures designs

Cross-over Designs

- Important assumption: No carry-over effects
 - effect of treatment received in each period is not affected by treatment received in previous periods.
- To minimize possibility of carry-over effects
 - “wash-out” time between the periods in which treatments are received.

Cross-over designs: Example

- Cross-over was done in actual experiment
- Each of 12 subjects observed under each condition
- Randomize order.
- One week period between observations.

Cross-over designs: Example

- Illustrates the importance of
 - “wash-out period” and
 - randomizing/balancing the order that treatments are applied.

In general, which design?

- Is the natural variability within a subject likely to be small relative to the natural variability across subjects?
 - More similarity within individuals or between individuals?
- Are there likely to be carry-over effects?
- Are there likely to be "drop-outs"?
- Is a cross-over design feasible?

Which design?

- No definitive statistical answer to the question.
- Answer depends on knowledge of
 - experimental material and
 - the treatments to be studied

Structure on the treatments: Factorial designs

- Example has four treatments: No water, Water, Isotonic G-E, Hypotonic G-E
- In other examples, in any of the designs we've considered, treatments can have *factorial structure*
 - Def: Treatments consist of combinations of factors

Change to a hypothetical example

- Suppose we had four treatments: No water, Water, G-E only, Water + G-E
- Combinations of factors: 1) Water 2) G-E

<u>Treatment</u>	<u>Water</u>	<u>G-E</u>
“No water”	Absent	Absent
“Water only”	Present	Absent
“G-E only”	Absent	Present
“Water + G-E”	Present	Present

Factorial designs

- Example is a “2 x 2 factorial” : two factors each at 2 levels
- Factorials can be done with any number of factors at any number of levels:
 - 2 x 2 x 2: three factors each at 2 levels
 - 3 x 4 2 factors: one at 3 levels, 1 at 4 levels

Factorial designs and statistical interaction

- To simplify, assume we do a completely randomized design with 24 subjects
 - 6 randomly assigned to each of 4 treatments
- Def: Two factors are said to “interact” if the effect of changing the level of one factor depends on the level of the other factor

Illustration of definition of statistical interaction

Question: Is

- the effect of adding G-E (I.e, changing level from absent to present) when *no water is given*

different than

- the effect of adding G-E (I.e, changing level from absent to present) when *water is given*
- If yes, there is statistical interaction
- If no, then there is no statistical interaction

Why is interaction important?

Estimating the G-E effect

<u>Treatment</u>	<u>Water</u>	<u>G-E</u>	<u>n</u>
1. "No water"	Absent	Absent	6
2. "Water only"	Present	Absent	6
3. "G-E only"	Absent	Present	6
4. "Water + G-E"	Present	Present	6

Why is interaction important?

Estimating the G-E effect

- If no interaction: estimate effect of G-E by avg of groups 3 & 4 - avg of groups 1 & 2
- Uses 12 subjects with G-E compared to 12 subjects not given G-E.
- Same number of subjects as if had
 - decided to give all subjects no water (or all water)
 - done a two-treatment experiment (G-E vs no G-E)

Why is interaction important?

Estimating the water effect

- If no interaction: estimate effect of water by avg of groups 2 & 4 - avg of groups 1 & 3
- Uses 12 subjects with water compared to 12 subjects not given water.
- Same number of subjects as if had
 - decided to give all subjects no G-E (or all G-E)
 - done a two-treatment experiment (water vs no water)

Why is interaction important? 2-for-1 experiment

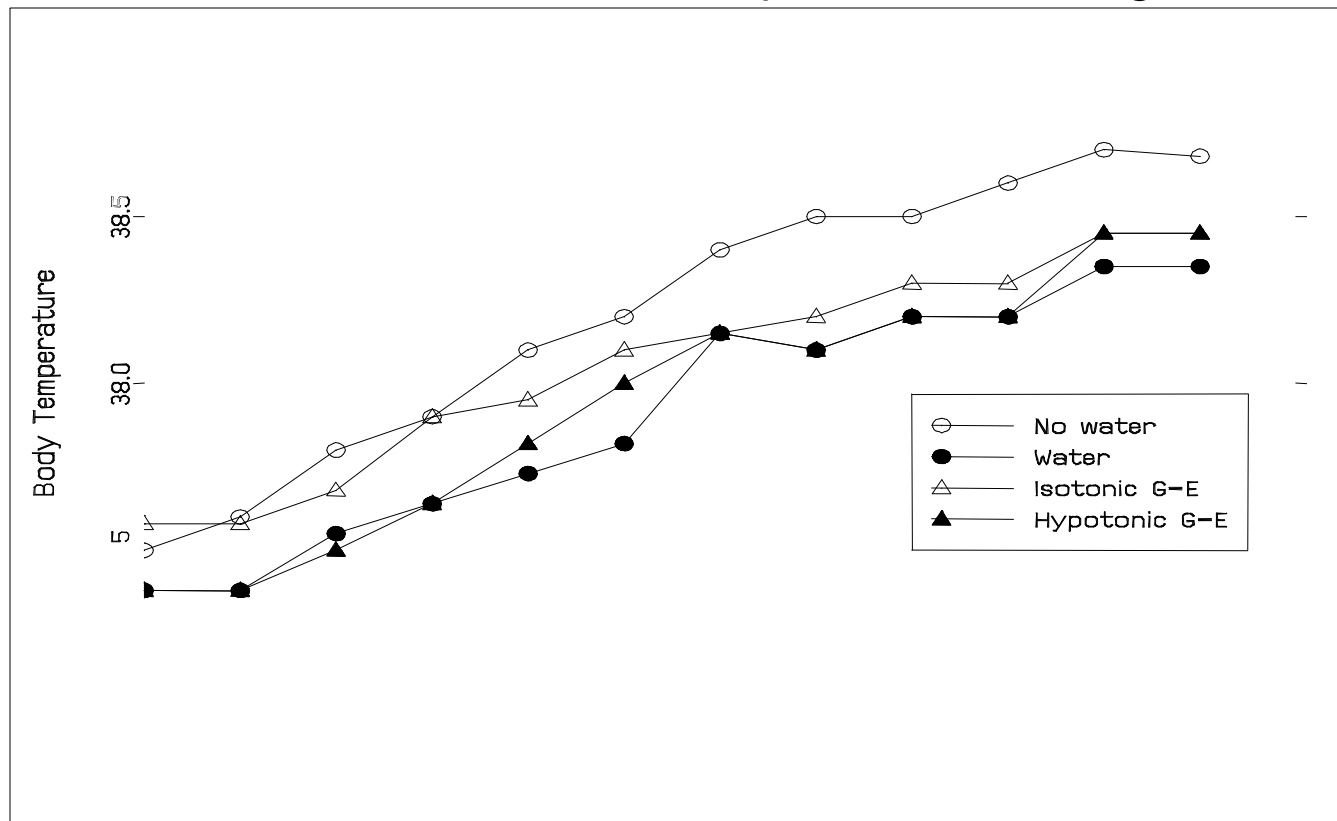
- If no interaction:
 - Get same “information” from 24 subjects as if had done 2 separate experiments, *each* with 24 subjects
- If there is interaction:
 - Hypothetical: May be important info: G-E not effective if water given but very effective if no water...
 - Best design for discovering it is a factorial

Serial measurements

- Observations taken repeatedly on same unit over time
- Can be done with any of the designs we've discussed
- Good overview given in Matthews et al.
Analysis of serial measurements in medical research (see letter to editor by S. Senn in same issue concerning this paper).
British Medical Journal, 300:230--235, 1990.

Example

- Example: Maughan et al (1996) take body temperature measurements over time while subject is exercising



Serial measurements

- Analysis should
 - take within-subject correlations into account or
 - be based on a summary measure
- Analyses generally should not
 - be done by comparing groups time point by time point

Serial measurements

- Important to consider within-subject profiles as well as trends across subjects.
- Otherwise
 - can be mislead as to the amount of variation or
 - the direction of effects

Handling dropouts in longitudinal studies

- Possible approaches.
- Analyze only those who complete therapy.
 - May bias results, especially if reason for dropout is related to outcome

Handling dropouts in longitudinal studies

- Use "Last Observation Carried Forward (LOCF)" method.
 - After patient has withdrawn, use the last observation.
 - Could bias results; last observation may not reflect true state of subject
 - Does not provide reasonable assessment of uncertainty
 - Generally dismissed as a method for handling dropouts

Handling dropouts in longitudinal studies

- Modeling the dropout process
 - Requires assumptions and sophisticated modeling methods.
- No generally accepted method for handling dropouts.