

A Primer on Health Outcomes Data Analysis

David M. Radosevich Ph.D., RN
University of Minnesota
davidmr@umn.edu

April 3, 2013

Topics

1. Analysis Appropriate to the Audience and Question
2. Observational Data and Salient Threats to Validity of Observational Studies
3. Character of the Outcome and the Types of Analyses

Analysis Appropriate to the Audience and the Question

- Health outcomes research starts with asking the question
 - Vaguely posed question... any analysis is satisfactory
 - Analysis pleases the analyst...but is a foil to the user
 - Good questions do not begin by identifying the source of data or an analytical method
 - Recognize the conundrum of using existing (e.g. records, clinical systems, registries, data marts) – could lack necessary data
- Have a model...
 - Conceptual model (confirmatory)
 - Operating model that specifies the variables and their interrelationship (exploratory)
- Development of the question and model are necessary antecedents to choosing an analytical strategy

Observational Studies

- Quasi-experiments lack random assignment to treatment
- Threats to the validity often ignored in conducting health outcomes data analyses
 - Selection
 - Low statistical power
 - Unreliability of treatment implementation
 - Inflated Type 1 error rate
 - Fishing and error rate problems
 - Statistical regression
 - Missing data

Selection

Differences between the kinds of people in the experimental group as opposed to control group

Examples:

- Differences in health (e.g. severity of disease) accounts for posttest differences between treatment and control (or comparison) groups
- Differences in subjects receiving one treatment versus another

Solutions:

- Risk adjustment for factors that differ between treatment and control groups and are associated with the outcome
- Randomization
- Matched designs
- Propensity scores

Propensity Score Methods

Background – in observational studies, investigator has no control over treatment assignment, leading to biased estimates of treatment effects.

Propensity score – the conditional probability of being treated given the subject’s covariates is used to balance the covariates in the groups. Adjusts the estimate of the treatment effect, creating a “quasi-randomized” experiment.

Common techniques – matching, stratification, and regression adjustment.

Example – Study to Understand Prognosis and Preferences for Outcomes and Preferences for Outcomes and Risks of Treatments (SUPPORT) found that right heart catheterization (RHC) increased mortality and increased utilization of services.

Low Statistical Power

Failure to detect hypothesized difference or likelihood of failing to reject the null hypothesis

Examples:

- Inadequate effect size for the outcomes being investigated
- Insufficient sample size or infrequent events

Solutions:

- *a priori* statistical power analysis
- Continuous variables > categorical variables
- Don't do the study

Unreliability of Treatment Implementation

Lack of standardized implementation of the treatment or differences in the way treatment is implemented

Examples:

- Aphorism – many observational studies are implemented without standardizing the treatment
- “Feel good” interventions
- Ignoring – failures to get the intervention fully implemented, crossing over to get another treatment, treatment diffusion
- Patient attrition

Solutions:

- Measure components of treatment and their relationship to outcomes
- Intent-to-treat analysis and analysis by the amount to treatment received

Inflated Type 1 Error Rate

Failure to account for all the sources of variability

Examples:

- Using traditional statistical tests (e.g., least squares linear regression) for repeated measures designs
- Treating replicate outcomes measures as independent

Solutions:

- Use mixed model methods to account for within subjects variability
- Account for inter-agency variability

Fishing and Error Rate Problems

Data dredging or multiple comparison problem

Examples:

- Failure to define *a priori* hypothesis leads to data dredging for statistically significant
- Use of multi-dimensional scales (e.g., SF-36 Health Survey) without first specifying the scales that have primary importance

Solutions:

- Statistical correction for multiple comparisons – Bonferroni, Scheffé
- Prepare primary and secondary hypotheses in advance of implementing the study

Statistical Regression

The scores for people selected as extreme (high or low) scorers regress toward the mean (average) with re-measurement

Examples:

- Cases exhibiting poor behavior at time of admission can 'only' move toward improved behavior on follow-up
- High knowledge levels at baseline are more likely to look worse at follow-up

Solutions:

- Avoid selecting cases on the basis of extreme scores
- Reduce measurement error by selecting outcomes measures with high reliability
- Standardization of measurement and training
- Multiple measurements; multiple visits

Ambulatory Versus Office Blood Pressure

	Office DBP	Ambulatory DBP	Difference
< 90 mm Hg	77	78	-1
90–104 mm Hg	96	93	3
> 105 mm Hg	111	101	10

Missing Data

Either admission or discharge outcomes data are missing

Examples:

- Attrition or lost to follow-up
- "If there are any ways in which data can be missing, they will be."

Solutions:

- Good process control
- Missing data are positive information
- Utilize existing information - interpolation, mean substitution, last value carried forward (LVCF)

Character of the Outcome and Types of Analysis

- "To scale or not to scale that is the question" - use of dichotomous categorical variables. Indices, model-based scoring methods (Likert, Guttman)
- Categorical - logistic regression, generalized estimating equation (GEE), generalized linear mixed models (GLIMMIX)
- Continuous - general mixed model (MIXED)
- Time-to-event analysis - survival analysis, Kaplan-Meier methods, Cox proportional hazards regression

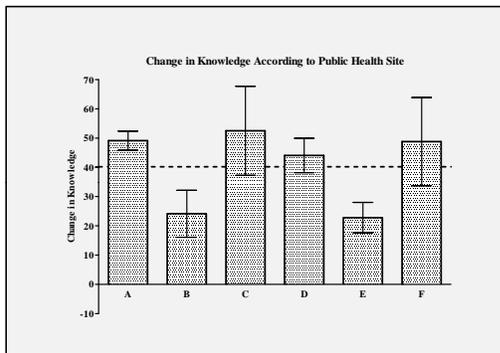
Real life examples...

...what interpretations can be made?

...what are threats to validity of inferences?

**Comparison of Change in Benchmark
for Knowledge by Public Health Site**

GLIMMIX Model

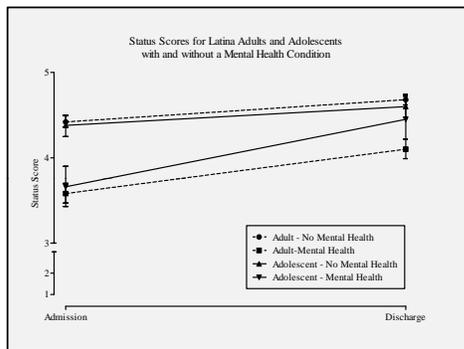


**Difference in Knowledge between Admission and Discharge
by Public Health Site**

	Admission	Discharge	Discharge - Admission	p value	Change Versus All Other Counties	p value
Site A	11.4 ± 1.5	60.6 ± 1.5	49.1 ± 1.7	< 0.001	10.7 ± 3.0	< 0.001
Site B	38.5 ± 3.8	62.6 ± 3.8	24.1 ± 4.1	< 0.001	-19.3 ± 4.7	< 0.001
Site C	23.0 ± 6.6	75.5 ± 6.6	52.5 ± 7.8	< 0.001	14.7 ± 8.0	0.067
Site D	26.5 ± 3.0	70.6 ± 3.0	44.1 ± 3.0	< 0.001	5.0 ± 3.8	0.038
Site E	21.3 ± 2.3	44.1 ± 2.3	22.8 ± 2.6	< 0.001	-20.9 ± 3.6	0.034
Site F	18.5 ± 6.5	67.3 ± 6.5	48.8 ± 7.7	< 0.001	10.3 ± 7.9	0.079

Comparison of Change in Status Score for Latina Adolescents with Mental Health Conditions

MIXED Model



Difference in Status between Admission and Discharge for Latina Adolescents and Adults with and without Mental Health Conditions

	Admission	Discharge	Change Score	Difference in Change	Difference of Differences	p value
No Mental Health						
Adolescent	4.38	4.60	0.22	0.04	0.31	0.012
Adult	4.42	4.68	0.26			
Mental Health						
Adolescent	3.66	4.45	0.79	-0.27		
Adult	3.58	4.10	0.52			

*Event Occurrence, Time-to-Event Analysis,
Survival Analysis*

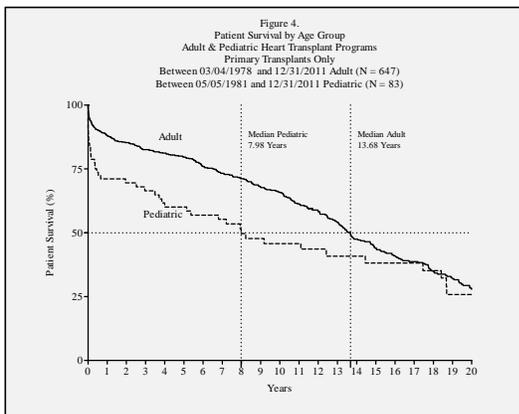
- The “whether” and “when” test (Singer)
- Framing a question
 - target event
 - beginning of time
 - metric for time
- Censoring
 - non-informative versus informative
 - right versus left censoring

22

Censoring

- Right censoring – terminate observations before everyone has experienced the event
- Left censoring – some have already experienced the event
- Random censoring – observations terminated for reasons not under the control of the investigator

23



Questions
