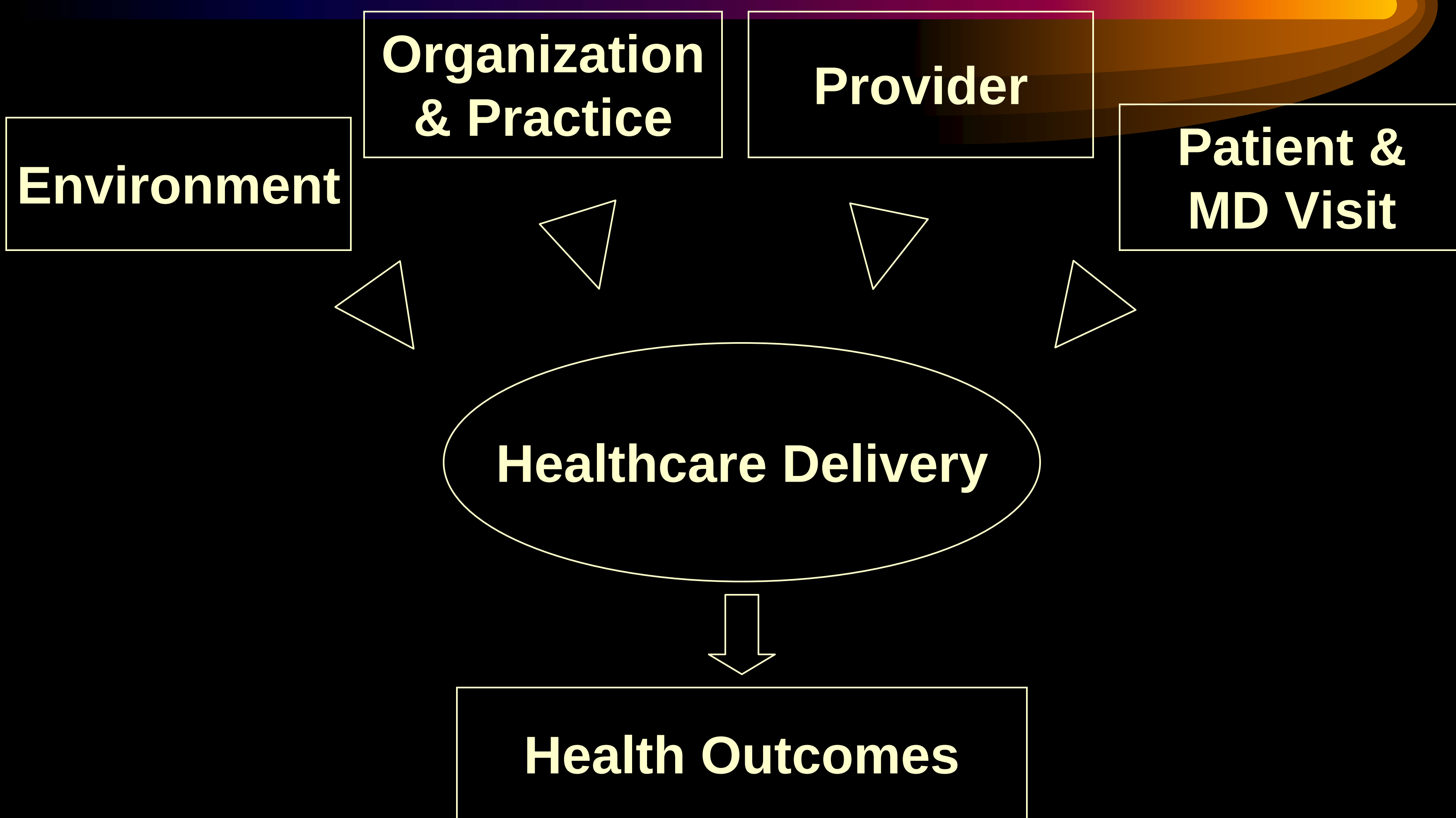
Hierarchical Research Designs: Design Strategies and Statistical Approaches

Martin Lee, PhD

Example: Influences on Delivery of Healthcare



Background

- Healthcare delivery interventions: Process of care
  - clinical guideline or clinical pathway implementation
  - collaborative care models
  - clinical care reorganization
  - managed care practice adoption (incentives, provider education, etc.)
  - require relatively complex research designs and sampling methods

Types of Interventions

- Act upon sometimes diverse provider groups and individuals
- Act upon the health care environment
- Frequently cross general medicine and specialty group lines of authority
- Require involvement of clinical and administrative leadership

Goals of Today's Discussion

- Review key research designs
  - eg, repeated measures, group randomization
- Review suitable sampling techniques
  - eg, hierarchical sampling of patients within providers within clinics
  - address issues of power and sample size
- Discuss statistical methods and software programs suited to these designs/samples

Process of Care Interventions Why Complex?

- Interventions implemented at provider level or higher (eg, clinic, hospital, hospital group, etc.)
- Outcome measured at lower level = hierarchical data
- Randomization of interventions at levels other than outcome

  Why?

  - "contamination" effect
  - difficulty in implementing multiple interventions in one facility
  - preferred technique in given facility
- Small number of groups per intervention

Examples

- **Rapid Early Action for Coronary Treatment (REACT) Trial (1995)**
  - goal to reduce time in seeking medical attention post-MI
  - intervention: includes mass-media campaign
  - randomize by city
- **Nutrition education**
  - goal to increase awareness of reduced fat/salt diet
  - intervention: label menus in restaurant
  - randomize by restaurant
- **QUITS Study**
  - goal to increase quit rates among veterans
  - intervention: EBQI guideline implementation
  - randomize by medical center

Design Issues

- **Group randomization** (similar to cluster or hierarchical random sampling): may not be able to view data as simple random samples
  - sampling unit ≠ analytical unit
  - unit of randomization ≠ unit of analysis
  - implications for sample size/power and analysis
  - "problem" clustering of data may not result from design
    - do we "test" for it? (conditional analysis)

Design Issues

- **Matching/stratification**
  - only a few (usually <10-20) group to be randomized
  - pair-matching?: difficult with only a few groups; usual advantage is gain in power:

    $\sigma^2_{diff} = 2\sigma^2_{within}(1 - \rho_m)$ (ρ_m = correlation between matching factor & outcome)

    with group-randomized, $\rho > 0.3$ or number of groups > 10 to be useful (Martin, 1993; Diehr, 1995)
  - stratification?: hard to detect stratum-by-outcome interaction
  - probably a good idea anyway because of small number of groups; match if close agreement is possible

Design Issues

- **Alternatives to matching/stratification**
  - post-hoc stratification (but stratification defined in advance)
  - regression adjustment for covariates (ANCOVA)

Design Issues

- **Repeated measures**
  - improve precision of within group data
  - t = 2 (“pre-test/post test” or panel design)
  - t > 2 (usual repeated measures design)
  - analytical issues

Sample Size Considerations

- Use "inflation" factor based on design effect of clustering (Donner & Klar, 2000)
- Fixed number of clusters/groups

Sample Size Considerations

- Easy method to employ
  - Inflation factor = $1 + (\bar{m} - 1)\,\rho$

    where $\bar{m}$ = average cluster size

    ρ = intracluster correlation coefficient

    $$= \frac{\text{variation between clusters}}{\text{variation between} + \text{variation within}}$$

  to traditional sample size formulae

- **Note significance of:**
  - $m = 1$ or $\rho = 0$
  - small ρ, but large m

Sample Size Considerations

How do we obtain these values?

$\overline{m}$ = average number of units to be sampled from each cluster (if this number varies widely, may need more exact approach using "m_i")

ρ = difficult to estimate; use survey information

INTRACLASS CORRELATIONS

EXAMPLES FROM PUBLISHED & UNPUBLISHED DATA

Published intraclass correlations:

General practice	**ICC**
• Fahey & Peters, BMJ 1996; 313:93-6 (prop pts controlled htn)	0.0644
• McKinley et al, BMJ 1997;314:190-3	
SF36 scores	
Physical functioning	0.00035
Role physical	<0.0001
Role emotional	0.019
Social functioning	<0.0001
Mental health	0.037
Energy and vitality	0.014
Pain	<0.0001
General health perception	<0.0001
Satisfaction scores	
Communication	0.056
Attitude of doctor	0.068
Continuity of care	0.019
Delay of visit	0.047
Overall satisfaction	0.058

UNPUBLISHED DATA

General practice	ICC
• North of England Study of Standards & Performance in Gen Pract (SPGP)	
Medical history recorded?	0.14
Child examined?	<0.01
Investigations recorded?	0.07
Previous diagnosis recorded?	<0.01
Diagnosis recorded?	<0.01
Previous drug management recorded?	<0.01
Drug management recorded?	<0.01
Non-drug management recorded?	0.08
Advice recorded?	0.11
Referral decision recorded?	<0.01
Follow-up decision recorded?	0.03
Reasons for management recorded?	0.16
Aberdeen Grampian Referral Initiatives Project (GRIP)	
Appropriateness of referral	0.04
Number of annual referrals	0.24
Aberdeen Urological Guideline Evaluation Project (URGE)	
SF36 scores	
Physical functioning	0.05
Role physical	0.01
Role emotional	0.008
Social functioning	0.02
Mental health	0.097
Energy and vitality	0.03
Pain	0.03
General health perception	0.02

Methodological Approaches for Analysis

<u>Example formula</u>:

- Comparison of two means (*no matching*)

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2 (2\sigma^2) [1 + (\overline{m} - 1)\rho]}{(\mu_1 - \mu_2)^2}$$

- Since $n = \overline{m}(k)$, then $k = n/\overline{m}$
  (# of groups)

- Comparison of two means (*matching*)

  $n_{match} = n\,(1 - \rho_M)$ (conservative: set $\rho_M = 0$)

Methodological Approaches to Analysis

- Efficiency of increasing number of clusters versus increasing number per cluster

$$\text{var}(\bar{y}_1 - \bar{y}_2) = \frac{2\sigma^2}{k\bar{m}}\left(1 + (\bar{m} - 1)\rho\right)$$

  - if k --> large, var --> 0
  - if m --> large, var --> $\dfrac{2\sigma^2\rho}{k}$

Methodological Approaches to Analysis

- Fixed number of clusters, k_{max}:

$$m = (1-\rho) \Big/ \left(\frac{k_{max}}{k_{m=1}} - \rho \right)$$

where $k_{m=1}$ = "usual" sample size (no clusters)

-- notice restriction on k_{max}:

$$k_{max} > \rho \, k_{m=1}$$

Methodological Approaches to Analysis

I. USE SUMMARY STATISTIC, e.g., mean, FOR ALL UNITS IN A GIVEN CLUSTER AND THEN UTILIZE TRADITIONAL STATISTICAL METHODOLOGIES ON THESE SUMMARY STATISTICS

Example: INTERVENTION #1 INTERVENTION #2

x clinic11	x clinic 21
•	•
•	•
•	•
x clinic n1	x clinic n2

Compare the means using t-test, say THIS MODEL ALLOWS FOR BASELINE CLUSTER-LEVEL COVARIATE ADJUSTMENT

BUT

NO AVAILABILITY OF SUBJECT LEVEL ADJUSTMENT i.e., information wasted

Methodological Approaches to Analysis

- **If two time points:**
  - use post-pre as measurement for analysis
  - better analyze post-data, adjusted for pre-data with ANOVA
  - better still, adjust individual data for covariates by regression model, then ANCOVA
  - even better, adjust by regression model, then repeated measures ANOVA
  - best?

Methodological Approaches to Analysis

II. INFLATE STANDARD ERROR OF USUAL TEST STATISTIC BY AN AMOUNT EQUAL TO

$$\sqrt{1 + (\bar{m} - 1)\,\rho}$$ (use $1 + (\bar{m} - 1)\rho$ for χ^2 and F-tests)

for 2-sample z or t-tests

$$t = \frac{x_1 - x_2}{S_w \sqrt{\frac{1}{n_1} - \frac{1}{n_2}}\,\sqrt{1 + (\bar{m} - 1)\,\rho}}$$

Where ρ is estimated as: $\dfrac{S^2_b}{S^2_b + S^2_w}$ Obtain from an ANOVA output

Problem: Tends to be conservative; examine *with* & *without* correction

EXAMPLE ON THE USE OF THE "CLUSTER-CORRECTED" T-TEST (Kish, 1965)

- examining differences in average income between homeowners and renters
- 40 neighborhoods (clusters)
- total sample of 400/group
- $\bar{m} = 40$
- estimate of intracluster correlation=.201

$$\text{So,} \quad \sqrt{1 + (\bar{m} - 1)\, r} \quad = \quad 2.97$$

Here $\quad \bar{x}_{\text{homeowner}} = \$40{,}000 \qquad \bar{x}_{\text{renter}} = \$35{,}000$

$S_w = \$16{,}000$

$$\text{So} \qquad t = \frac{40{,}000 - 35{,}000}{16{,}000 \sqrt{\dfrac{1}{400} + \dfrac{1}{400}}\ (2.97)}$$

$$= 1.49 \qquad (P=.14)$$

(UNINFLATED: T = 4.42, p < < .001!!)

Methodological Approaches to Analysis

III. USE MIXED MODEL ANALYSIS OF VARIANCE (quantitative data)

- Unit of analysis becomes repeat factor and cluster unit becomes a “nested” factor within “intervention” group
- Sometimes thought of as a random coefficient regression model
- Relationship between outcome variable and cluster may (and does) vary from cluster to cluster
- Can use with repeated measures/covariates

Methodological Approaches to Analysis

IV. HUBER CORRECTION TO STANDARD ERRORS

- Bootstrap/Jackknife approach (nonparametric)
- Used in connection with:
  - standard regression
  - logistic regression
  - Cox regression
  - Others
- Implemented in STATA

Methodological Approaches to Analysis

V. OTHERS (similar to repeated measures analysis)

- Panel data analysis
- MANOVA
- Growth curve models (like random coefficients)
- Bayesian/Empirical Bayesian

Software

- STATA (define cluster variable in regression procedures)
- SUDAAN (generalized estimation equations)
- WESREG (SAS procedure callable in v. 6.08 and higher)
- BMDP 3V/5V (repeated measures)
- SAS PROC MIXED (*not* PROC GLM; works only when data are balanced and covariates are "well-behaved")
- Others
  - GENMOD
  - HLM (2 or 3-level hierarchies)
  - ML3 (2 or 3-level hierarchies)
  - VARCL (up to 9-level hierarchies)

Example: *Kerr et al (1997)*

DOES DISSATISFACTION WITH ACCESS TO SPECIALISTS AFFECT THE DESIRE TO LEAVE A MANAGED CARE PLAN?

- Surveyed >120 physician groups throughout California (from particular health plan)
- Obtained 17,196 patients from within these groups
- Examined
  - Satisfaction scales (quantitative)
  - Desire to change health plan (binary)
- Huber regression using average number of enrollees per group

Example:

STUDY 0F THE RATE OF ADHERENCE TO TREATMENT REGIMENS IN VETERANS WITH HIV

- **Randomize clinics at regional level**
  - San Diego, Greater Los Angeles, Tucson and Palo Alto
- **Sample size calculation for t-test comparison inflated by a factor of about 1.5 to account for almost certain clustering effect**
- **Allowance for matching/stratification?**

Stepped Wedge Design

- Type of crossover design
- Different clusters crossover (1 direction) at different time points
- First time point: baseline
- Then, different clusters initiate the intervention at different time points (often by randomization mechanism

Stepped Wedge Design

Cluster	Time 1
1	1
2	1
3	0
4	0

Clus ter	Time 1	Time 2	Time 3	Time 4	Time 5
1	0	1	1	1	1
2	0	0	1	1	1
3	0	0	0	1	1
4	0	0	0	0	1

Cluster	Time 1	Time 2
1	1	0
2	1	0
3	0	1
4	0	1

Parallel group

X-over

Stepped Wedge

Stepped Wedge Design: Advantages

- Parallel/x-over designs: intervention implemented in half of all clusters simultaneously: may be logistically impossible; stepped wedge allows for limited rollout
- Only one direction for treatment intervention (not removed); however, complicates analysis (treatment effect cannot be estimated from within-cluster comparisons)

Example: Golden et al, NEJM, 2005

- Partner notification of patients with STDs
- Standard: public health authorities notify
- Intervention: patients given drugs/drug vouchers to give to partners
- Implemented in one county in WA and then randomly rolled out in multiple counties over time

Statistical Model

- Y_{ijk} = response corresponding to individual k at time j from cluster I
- $\mu_{ij} = \mu + \alpha_i + \beta_j + X_{ij}\theta$, where α_i is the random effect for cluster i, β_j is the fixed effect for time interval j, X_{ij} is an indictor variable for treatment and θ is the treatment effect
- $Y_{ijk} = \mu_{ij} + e_{ij}$
- Usual assumptions on the random effects

Statistical Analyses

- Linear mixed models
- Generalized linear mixed models (no normality
- Generalized estimating equations (allows for misspecification of the variance-covariance structure
- Research ongoing to compare these approaches