Clustered Measurement in Cluster Randomized Trials

Robert W. Platt

Department of Epidemiology, Biostatistics, and Occupational Health
Department of Pediatrics
McGill University

June 2009

Outline

Background
- Cluster RCTs
- Intra-cluster correlation
- Measurement in cluster RCTs

Case Study: PROBIT
- PROBIT: design and analysis
- Results of 6.5 yr follow-up

Model to "adjust" for clustered measurement
- Model

Conclusions and future work
- Conclusions

Cluster RCTs

- Clusters (hospitals, schools, communities) randomized to intervention/control
- Very useful for group-level interventions
- Useful for individual-level interventions when contamination is a problem
  - Contact between subjects randomized to different groups can create overlap and bias towards the null

Clustering

- Disadvantage: individuals within same group are correlated
  - Similar populations
  - Similar intervention delivery
  - Exposed to similar co-interventions
- Intra-cluster correlation (ICC) measures degree of correlation between cluster members
- ICC $\rho = \frac{\sigma_b^2}{\sigma^2+\sigma_b^2}$ where σ^2 is within cluster variance, σ_b^2 between cluster variance.

ICC

- $\rho = \frac{\sigma_b^2}{\sigma^2+\sigma_b^2}$
- $\rho = 0$ (i.e., $\sigma_b^2 = 0$) means no correlation (independent observations)
- $\rho = 1$ (i.e., $\sigma^2 = 0$) means complete correlation (within-cluster observations provide no additional info)
- ρ measures degree to which individual (within-cluster) observations represent independent observations

Measurement Issues

- In most studies, measurement is clustered
  - Observer - eg IQ measures
  - Instrument - eg blood pressure/poorly calibrated sphygmomanometers
- If inter-observer variability, will result in correlation between observations measured by same observer (i.e., $\sigma_b^2 > 0$, and $\rho > 0$

"Double Jeopardy"

- Cluster trials: often a single observer per cluster
- Often prohibitive (location, time) to have single observer for all clusters
- Problem: observer-driven clustering and inherent clustering are indistinguishable
- Additive effect of clustering can cause substantial problems

Design

- Randomized cluster trial of breastfeeding promotion intervention
- 31 hospitals in Belarus randomized to promotion or standard care, 17,046 subjects.
- Original outcomes: infection, growth in first year of life.
- Significant effect on GI infection; trend for respiratory infection
- Growth similar in intervention vs. control group

Analyses

- Originally matched by centre
  - Analysis to be based on pair-matched centres
- Data problems, dropouts (two centres) broke matching
- Revised analysis plan: linear/generalized linear mixed models accounting for clustering by hospital
  - And repeated measures by subject (for some outcomes)
- Adjusted for cluster-level covariates

PROBIT - "Audit" data

- Audit done for QC/validation purposes
- Aside: audit detected significant data problems at one site; site excluded from further analyses
- Design
  - Small number of subjects per site (5-10)
  - Single observer across ALL sites

Long-term follow up

- Children followed to age 6.5 ($\pm$ 0.5)
- Outcomes: IQ, school performance, behaviour, anthropometry, allergy, dental health
- Significant effect on verbal IQ; null on most other outcomes
- Some ICCs MUCH higher than expected (0.2-0.3 vs. 0.01-0.05).
- Substantial impact on statistical power

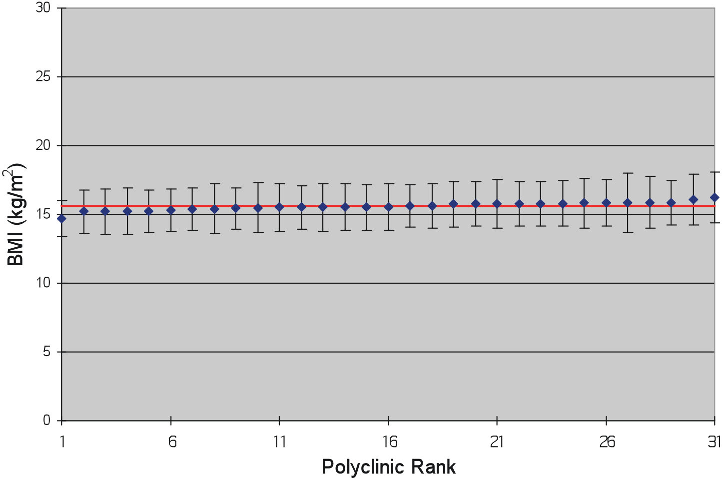
Future follow up

- Pre-puberty (9-11 yrs old) follow up underway
- Anthropometry, blood draws, metabolic syndrome.

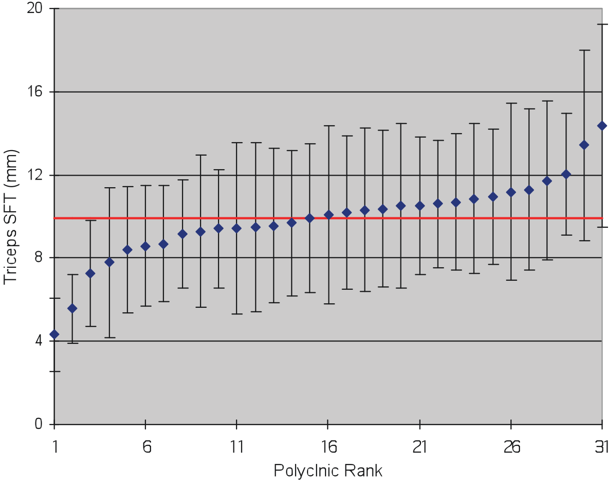
IQ measurement

- Single observer (trained pediatrician) assessed all children at each site
- Pediatricians were trained and calibrated by a central team
- Didn't work...

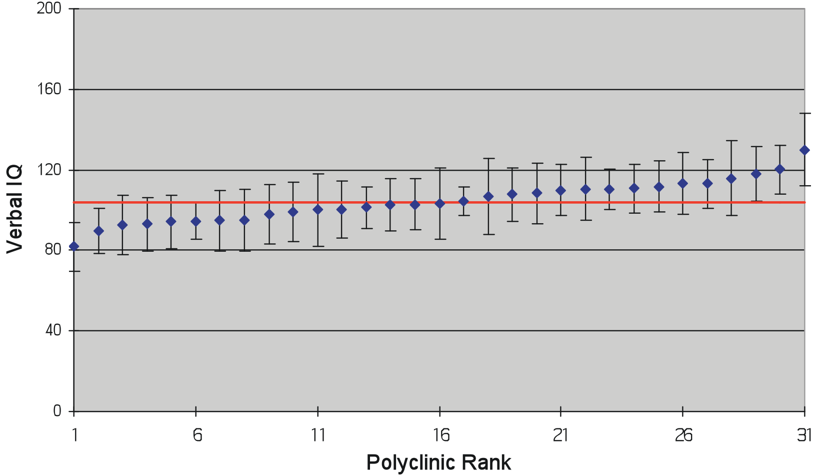
BMI - almost no clustering?



Triceps skinfold - some clustering



IQ measurement - evidence of clustering



Selected results at 6.5 yrs

ITT analysis

	Mean			
Outcome (Units)	Exp	Ctl	ICC	Adj Δ (95% CI)
BMI (kg/m2)	15.6	15.6	0.03	0.1 (-0.2, 0.3)
Tr skinfold (mm)	9.9	10.0	0.18	-0.4 (-1.8, 1.0)
Verbal IQ	108.7	98.7	0.31	7.5 (0.8, 14.3)

- Note: ICC varies substantially
- ICC of 0.31 implausible based on expected socioeconomic/cultural/genetic clustering
- ICC appears to be correlated with degree of observer influence

Random effects models for clustered data

$$Y_{ij} = \beta_0 + \beta_1 * T + b_i + \epsilon_{ij} \tag{1}$$

where

- $T = 1$ for the intervention group, 0 for the control group,
- b_i and ϵ_{ij} error terms at the cluster and individual level.
- Assume b_i and ϵ_{ij} normally distributed and independent of each other, with variances σ_b^2 and σ^2

$$\rho = \frac{\sigma_b^2}{\sigma^2 + \sigma_b^2}, \tag{2}$$

Clustered measurement

$$Y_{ij} = \beta_0 + \beta_1 * T + b_i + d_i + \epsilon_{ij} \quad (3)$$

where

- d_i random effect due to measurement (variance σ_d^2)
- b_i and d_i are not directly identifiable at the individual level.
- Assume b_i and ϵ_{ij} normally distributed and independent of each other, with variances σ_b^2 and σ^2

$$\rho_c = \frac{\sigma_b^2 + \sigma_d^2}{\sigma^2 + \sigma_b^2 + \sigma_d^2}. \quad (4)$$

Problem...

- Observed data generated by equation 3
- but what if n_i subjects per cluster, $n_i \leq N_i$, are measured by a single auditor for all sites?
- And we assume these audited measures come from equation 1 (i.e, $\sigma_d^2 = 0$)

$$Y_{ijA} = \beta_0 + \beta_1 * T + b_i + \epsilon_{ij} \tag{5}$$

  - This was done in PROBIT!

Solution

Bayesian inference

- Problem is akin to missing data
- Bayesian methods (using WinBUGS) easily implemented for these problems
- Set up: n_i subjects per cluster have both Y_{ij} and Y_{ijA}; $N_i - n_i$ have Y_{ij} only.
- For the latter, Y_{ijA} is a missing value

Inference - Details

- Diffuse priors on all parameters
- Model relatively straightforward to fit

IQ Results - Summary

Parameter	Mean	SD	Credible Int.
β_0	101.183	2.097	(96.88, 104.783)
β_1	5.417	3.243	(1.121,11.19)
σ^2	202.0	2.404	(197.3, 206.7)
σ_b^2	51.0	20.19	(20.88, 98.89)
σ_d^2	43.36	18.19	(16.08, 86.46)

- $\hat{\rho}_c = 0.31$, $\hat{\rho} = 0.20$
- About half of the ICC is due to clustered measurement!
- Main effect - adjusted result is shifted towards null, but 95% CI narrower.

Conclusions

- In cluster RCTs clustered measurement matters
- Calibration of observers across sites is *critical*
- Multiple observers per site, or single observer for all sites, can help
- Having some observers test subjects at multiple sites can help
  - Allows identification of σ_b^2, σ_d^2
- Additional data can help

Future work

- How bad can it be, and how much can extra data help?
- Test other outcomes
- Theoretical results?
  - Assume n_i, σ_b^2, σ_d^2. Can we compute efficiency?
- Simulations?
- Design considerations - how to design the next cluster trial?
  - Efficient use of observers for a fixed cost, or minimize cost for a fixed variance

Thanks

- Michael Kramer and the PROBIT study investigators and funders
- Richard Martin
- Jonathan Sterne
- Stan Shapiro