

# **Standalone use of “STATA” for analysis of cluster randomized controlled trials (cluster RCT)**

**Dr Sahul Bharti**

*MD Pediatrics (PGI, Chandigarh)*

*Pediatric Endocrinologist (SGPGI, Lucknow)*

**Founder**

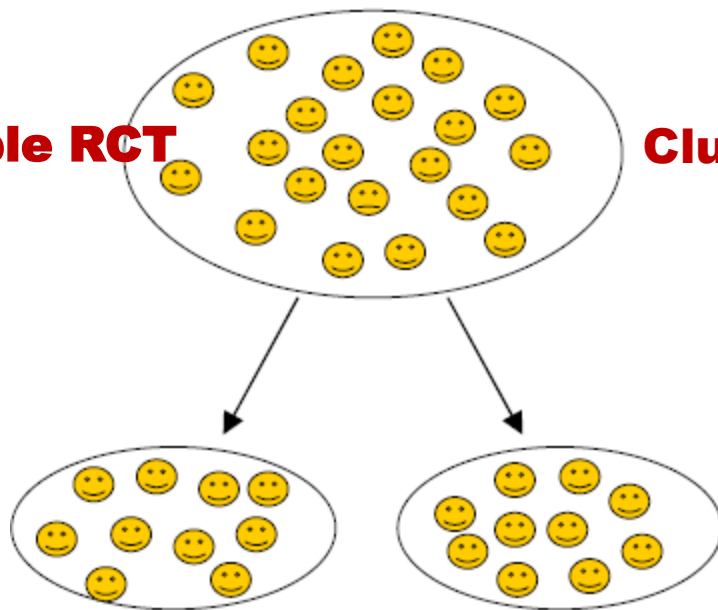
**Build Healthy India Movement (BHIM)  
(Research based NGO)**

# Cluster Randomized Trials

... are clinical trials (experiments) in which **social units** or **clusters of individuals** rather than independent individuals are **randomly allocated** to intervention groups.

Individual randomization

**Simple RCT**

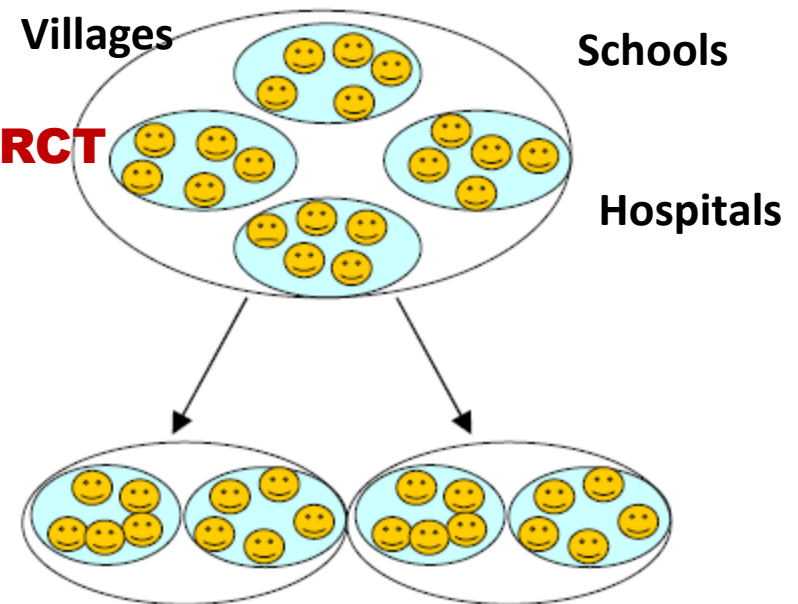


Intervention group

Control group

Cluster randomization

**Cluster RCT**



Intervention group

Control group

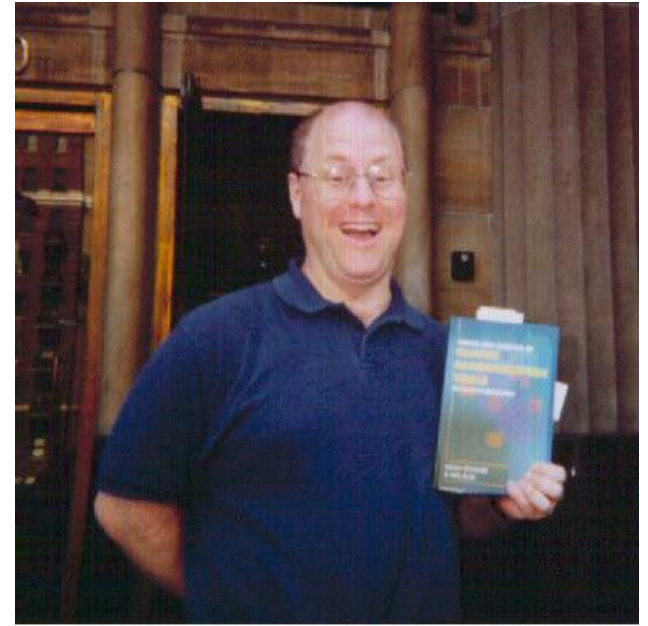
# Pioneers on “group or cluster RCT”



**David M Murray**



**Dr. Allan Donner**



**Dr. Neil Klar**

- *A Community-Based Cluster Randomized Controlled Trial of "Directly Observed Home-Based Daily Iron Therapy" in Lowering Prevalence of Anemia in Rural Women and Adolescent Girls.*

*Bharti S, Bharti B et al* Asia Pac J Public Health. 2013 May 10

32 villages

Intervention in  
16 villages (524  
anemic women)

Trial was 90 days of treatment with  
iron tablets

Control in  
16 villages (535  
anemic women)

Directly observed  
home based iron  
therapy

*Two key outcomes on follow up:*  
• **Decrease in anemia prevalence**  
• **Increase in mean Hb values**

Unsupervised clinic  
driven therapy

*On follow up after 3 months*

16.8% anemia

Mean Hb  
13.01 gm/dl

35.5% anemia

Mean Hb  
12.32 gm/dl

**Sample  
size**

**“clustersampsi”**

**ICC**

- Quantitative outcome (loneway)
- Binary outcome (loneway)

**Cluster  
level  
analysis**

- Parametric (t-test)
- Non parametric (ranksum)

***Resampling (Permutation tests)***

**Individual  
level  
analysis**

- Standard regression with cluster adj. robust SE
- Random effect/mixed effect modeling
- Generalized estimated equations (gee)

# Intraclass correlation coefficient

$$ICC = \frac{\text{Between cluster variability}}{\text{Within cluster variability} + \text{Between cluster variability}}$$



***ICC=0***



***ICC=1***

# ICC for quantitative outcome

## loneway Hb village

loneway - Large one-way ANOVA

Main by/if/in Weights

Response variable: Hbinitial  
Group variable: Village

Reference point for estimation of rho

Default value of 1  
 Expected value of F distribution  
 Median of F distribution

Exact confidence intervals (groups must be equal)

95 Confidence level

OK Cancel Submit

Intraclass correlation	Asy. S.E.	[95% Conf. Interval]	
0.02741	0.01434	0.00000	0.05552
Estimated SD of villagecode effect		.1814422	
Estimated SD within villagecode		1.080792	
Est. reliability of a villagecode mean		0.48194	
(evaluated at n=33.01)			



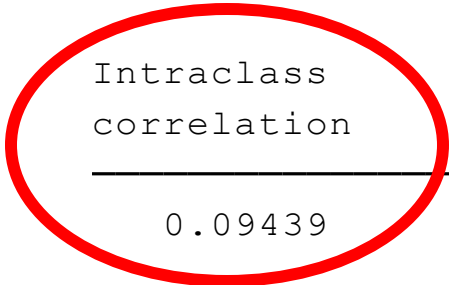
# ICC for qualitative outcome

## loneway anemia village

One-way Analysis of Variance for AnemiaFU: FU Anemia prevalence

Number of obs = 1059  
R-squared = 0.1183

Source	SS	df	MS	F	Prob > F
Between villagecode	24.191935	31	.780385	4.44	0.0000
Within villagecode	180.35386	1027	.17561233		
Total	204.5458	1058	.19333251		



Intraclass correlation	Asy. S.E.	[95% Conf. Interval]	
0.09439	0.02884	0.03787	0.15092

Estimated SD of villagecode effect .1352934  
Estimated SD within villagecode .4190612  
Est. reliability of a villagecode mean 0.77497  
(evaluated at n=33.04)

# Sample size in cluster or group RCT

**Simple RCT:**  
Sample size per arm= 121  
Total sample size= 242

ssi 0.50 0.30, alpha(0.05) power(0.9)

Sample  
size

**Stata**

**m**

Av cluster size

**$\rho$**

Intracluster correlation  
Coefficient (ICC)

**$\rho$**

**$\alpha$   $\beta$**

Alpha & Beta errors

**P1-P2**

Minimal detectable  
difference

**Cluster RCT:**  
Sample size per arm= 480  
Number clusters per arm=16

**clustersampsi**

clustersampsi, binomial samplesize p1(.50) p2(.30) m(30) rho(0.05) alpha(0.05) beta(0.90)

# Dialog Box (Menu driven) for sample size in cluster RCT

Sample Size Calculations for Cluster RCTs

Main Options Values

Calculation Type

- Compute Sample Size
  - Specifying Average Cluster Size
  - Specifying Number of Clusters
- Compute Power
- Compute Detectable Difference

Cluster Heterogeneity

- Specify via ICC  
ICC .05
- Specify via CV  
CV

Data Type

- Two sample comparison of means
- Two sample comparison of proportions
- Two sample comparison of rates

? P [document icon]

OK Cancel Submit

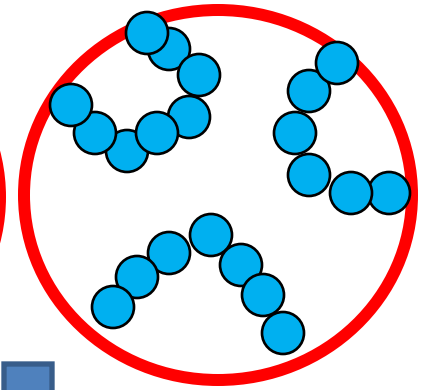
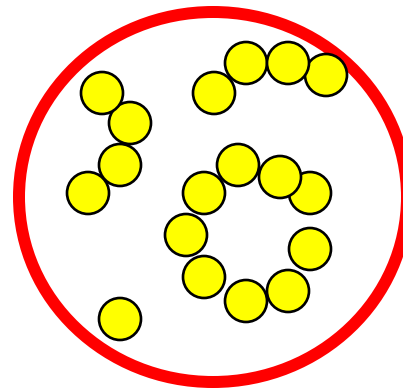
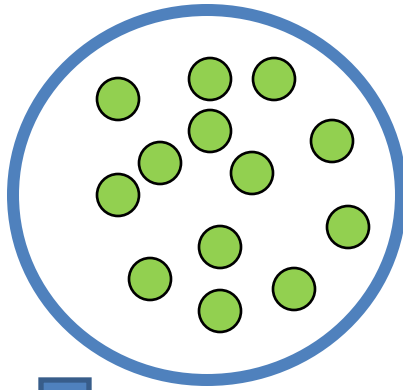
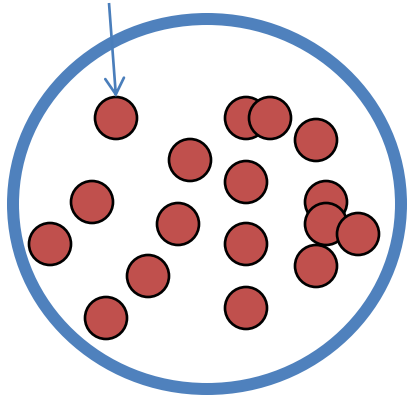
# Statistical analysis of cluster RCT outcomes

## *Two key approaches:*

- *Cluster level analysis by using summary measures*
- *Individual level analysis adjusted for clustering*

# Cluster level analysis

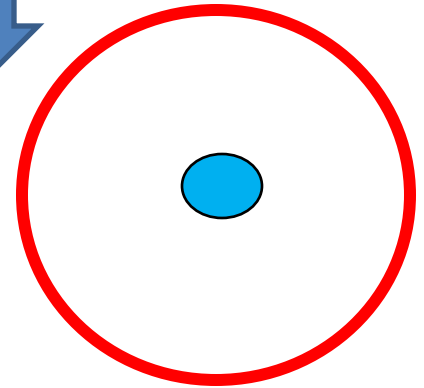
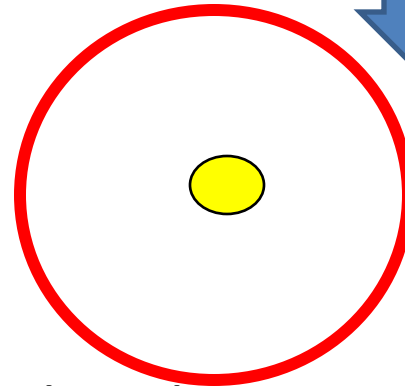
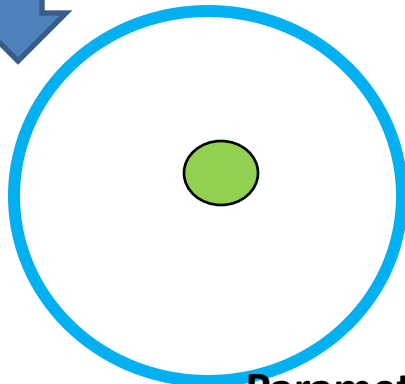
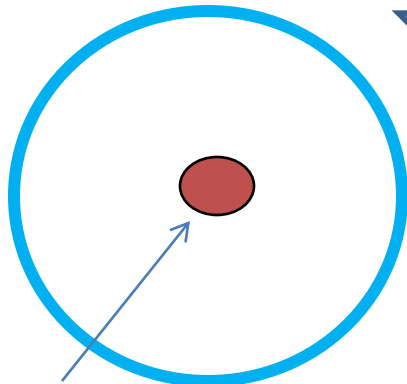
*Individual response*



**collapse (mean) Hb, by (cluster)**

**Intervention clusters**

**Control clusters**



Summary value

**Parametric tests ( t-test)**



**Non-Parametric tests ( Ranksum test)**



**Resampling tests ( Permutation test)**



# Permutation test

```
permute HB d=(r(mu_1) - r(mu_2)), reps(5000) : ttest HB, by ( Studygroups )
```

Monte Carlo permutation results

Number of obs = 32

```
command: ttest HBfu, by( Studygroups )
```

```
d: r(mu_1) - r(mu_2)
```

```
permute var: HBfu
```

T	T(obs)	c	n	p=c/n	SE(p)	[95% Conf. Interval]
d	-.6911765	1	5000	0.0002	0.0002	5.06e-06 .0011138

Note: confidence interval is with respect to  $p=c/n$ .

Note:  $c = \#\{|T| \geq |T(\text{obs})|\}$

**Individual level  
analysis adjusting for  
clustering and  
covariates**

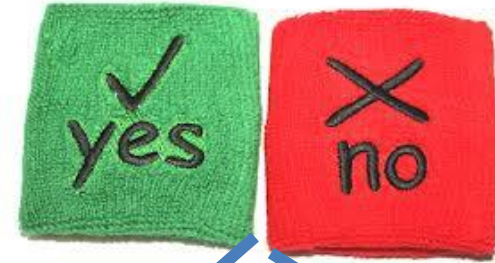
# Analysis of binary variable

Chi-square test:  
`tab Anemia Studygroups , chi2`

Cluster adjusted Chi-square test:  
`clchi2 Anemia Studygroups , cluster( village)`



	Good/Yes	Bad/No
Intervention	%	%
Control	%	%



**Binary logistic Regression (logit)**

**Cluster specific approach:**  
Random/mixed effect logistic Regression  
(`xtlogit`/`xtmelogit`)

**Population averaged approach:**  
Generalized estimating equations  
(`xtgee`)

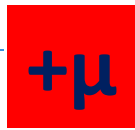


Standard logistic regression:

logit Anemia Study groups Age Hb baseline

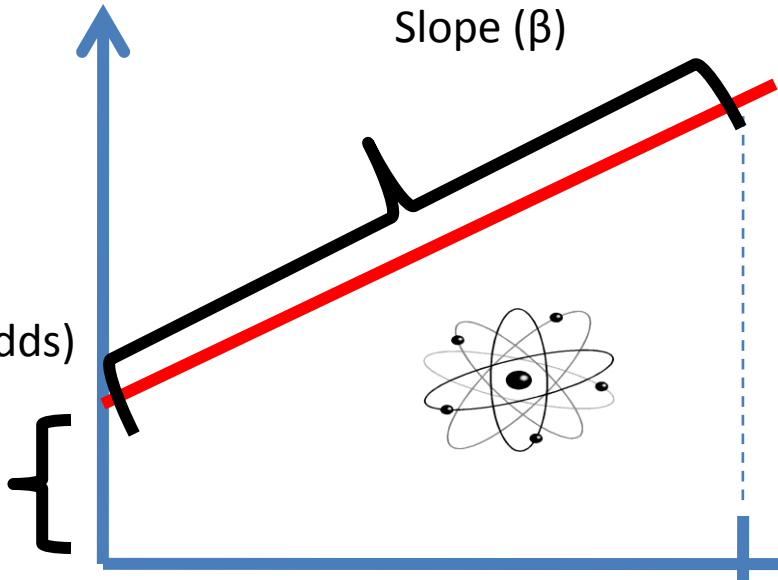
$$\text{Log odds of } Y = \alpha + \sum_1^j \beta X + \mu$$

Slope ( $\beta$ )



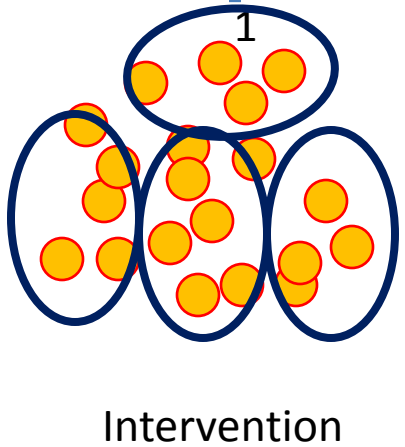
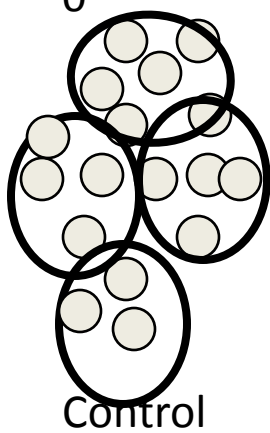
Probability of Anemia (log odds)

Intercept ( $\alpha$ )



Random effect log regression:  
xtlogit Anemia Study groups Age Hbinitial, re (village)

Generalized estimating equation:  
xtgee Anemia Study groups Hbinitial Age , family(poisson) link(log) corr(exchangeable) vce(robust)



Control

Intervention

```
. logistic AnemiaFU Studygroups Reproductive Hbinitial Age, vce(cluster villagecode)
. xtlogit AnemiaFU Studvargroups Reproductive Hbinitial Age, re or
```

```
Random-effects logistic regression
Group variable: villagecode

Random effects u_i ~ Gaussian

Number of obs      =      1023
Number of groups   =         32

Obs per group: min =         16
                avg  =        32.0
                max  =         51

Wald chi2(4)      =        56.68
Prob > chi2       =        0.0000

Log likelihood    =   -535.0275
```

AnemiaFU	OR	Std. Err.	z	P> z	[95% Conf. Interval]	
Studygroups	.3283787	.0793879	-4.61	0.000	.2044518	.5274229
Reproductive	1.145492	.1899261	0.82	0.413	.8276761	1.585346
Hbinitial	.6744372	.0451968	-5.88	0.000	.5914241	.7691021
Age	.9925296	.007035	-1.06	0.290	.9788365	1.006414
_cons	37.15257	28.31388	4.74	0.000	8.342387	165.4579
/lnsig2u	-1.379775	.4716034			-2.3041	-.4554488
sigma_u	.5016326	.1182858			.3159883	.7963437
rho	.0710532	.031128			.0294563	.1616101

```
Likelihood-ratio test of rho=0: chibar2(01) = 15.22 Prob >= chibar2 = 0.000
```

xtgee AnemiaFU Studygroups Reproductive Hbinitial Age, family(poisson) link(log) corr(exchangeable) vce(robust) eform

xtgee - Fit population-averaged panel-data models by using GEE

Model | Model 2 | Correlation | by/f/in | Weights | SE/Robust | Reporting | Optimization

Dependent variable: AnemiaFU  
Independent variables: Studygroups Reproductive Hbinitial Age

Panel settings...

Family and link choices:	Gaussian	Inverse Gaussian	Binomial	Poisson	Negative binomial	Gamma
Identity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Log	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Logit			<input type="radio"/>			
Probit			<input type="radio"/>			
C. log-log			<input type="radio"/>			
Power	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Odds power			<input type="radio"/>			
Neg. binom.					<input type="radio"/>	
Reciprocal	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

OK Cancel Submit

```

GEE population-averaged model          Number of obs      =      1023
Group variable:          villagecode    Number of groups    =       32
Link:                    log            Obs per group: min =       16
Family:                  Poisson        avg =              32.0
Correlation:            exchangeable    max =              51
                                                Wald chi2(4)       =      81.67
Scale parameter:        1              Prob > chi2        =      0.0000

```

(Std. Err. adjusted for clustering on villagecode)

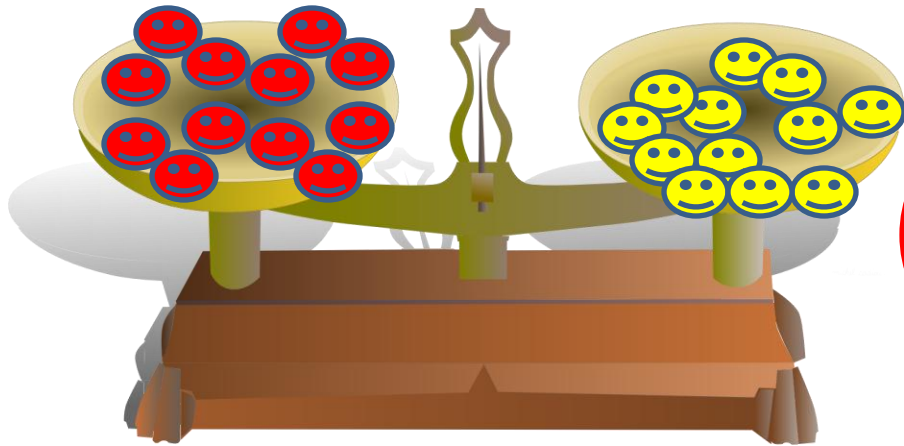
AnemiaFU	Robust					
	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
Studygroups	.4673556	.079564	-4.47	0.000	.3347622	.6524669
Reproductive	1.08832	.109277	0.84	0.399	.893899	1.325027
Hbinitial	.7974345	.0262511	-6.88	0.000	.7476081	.8505818
Age	.995109	.0063987	-0.76	0.446	.9826464	1.00773
_cons	4.004446	1.372428	4.05	0.000	2.045566	7.839193

# Analysis of a quantitative variable

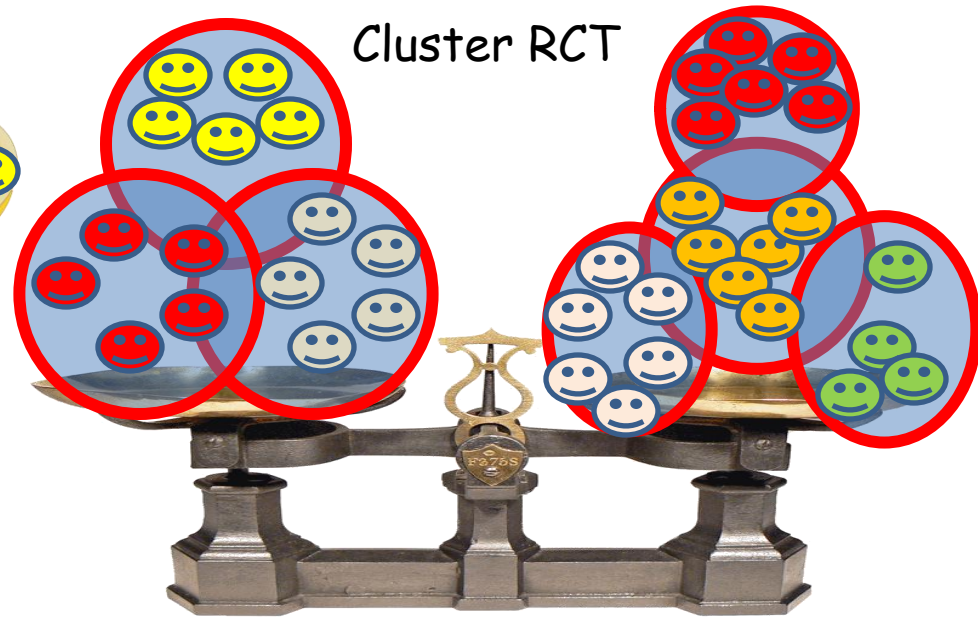
Univariate: t-test  
ttest HB, by(Study groups)

Univariate: cluster adjusted t-test  
clttest HB, cluster( village) by(Studygrps )

Simple RCT



Cluster RCT



Multivariate: Linear Regression (regress)

Linear Regression with clustered robust SE  
Random effect linear Regression (xtreg)  
Mixed effect linear Regression (xtmixed)  
Generalized estimating equations (xtgee)

```
. regress HBfu Studygroups Reproductive Hbinitial Age, vce(cluster villagecode)
. xtreg HBfu Studygroups Reproductive Hbinitial Age, re vce(cluster villagecode)
```

```
Random-effects GLS regression           Number of obs       =       1023
Group variable: villagecode            Number of groups     =         32

R-sq:  within = 0.1217                  Obs per group: min =         16
      between = 0.5309                               avg =        32.0
      overall  = 0.1809                               max =         51

                                           Wald chi2(4)        =        98.48
corr(u_i, X) = 0 (assumed)              Prob > chi2         =        0.0000
```

(Std. Err. adjusted for 32 clusters in villagecode)

HBfu	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
Studygroups	.6896205	.1287775	5.36	0.000	.4372213	.9420198
Reproductive	-.0635847	.0650008	-0.98	0.328	-.190984	.0638146
Hbinitial	.4074096	.0567077	7.18	0.000	.2962644	.5185547
Age	.0026616	.0043214	0.62	0.538	-.0058081	.0111313
_cons	7.99799	.6351651	12.59	0.000	6.75309	9.242891
sigma_u	.30138664					
sigma_e	1.1975596					
rho	.05956389	(fraction of variance due to u_i)				



A large, glowing full moon in a dark sky, partially obscured by the silhouette of a castle on a mountain peak. The moon is bright orange and yellow, and the castle is dark and silhouetted against the moon and the sky. The background shows a mountain range with snow or light-colored rock.

**THANK  
YOU**