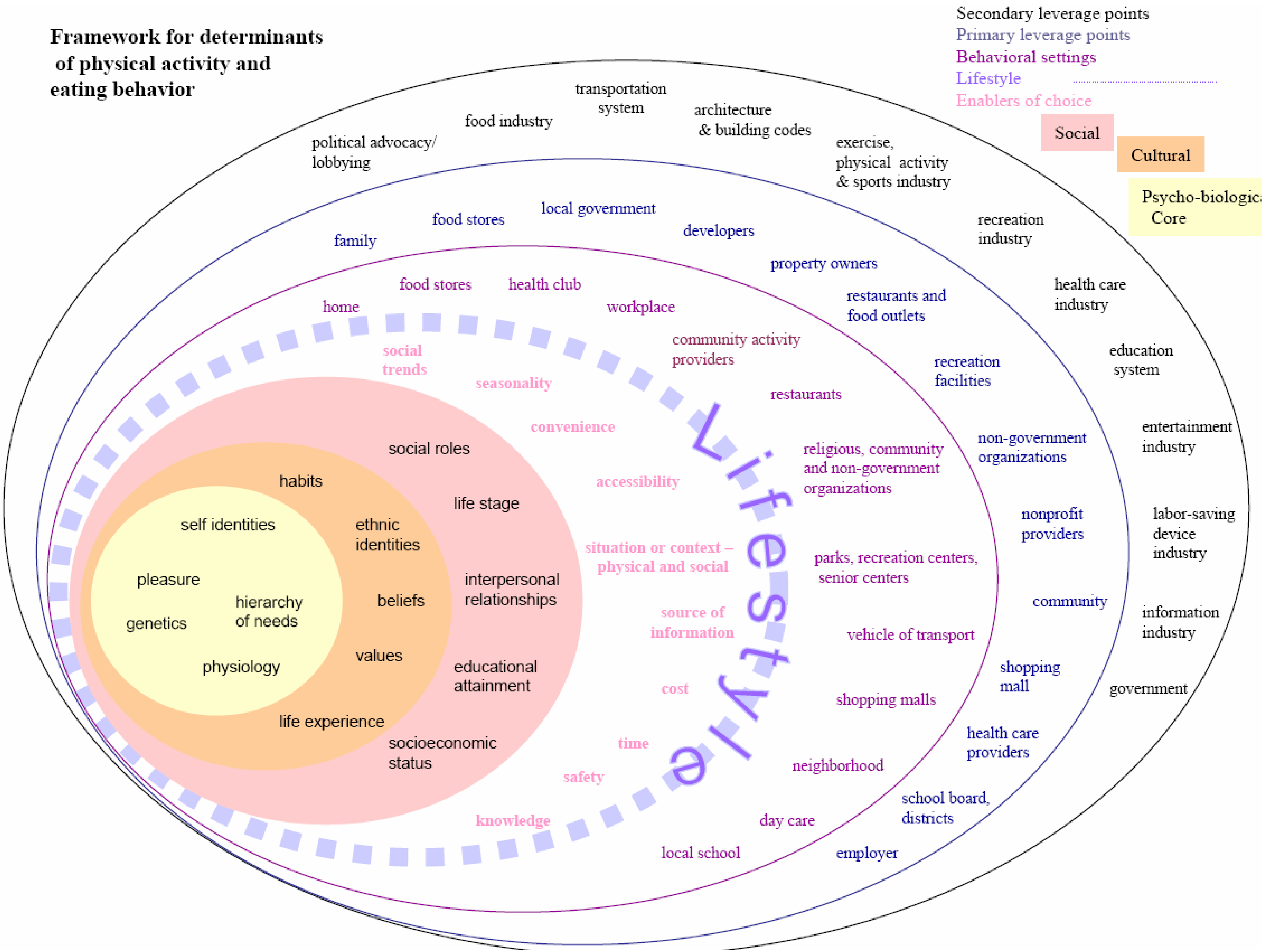
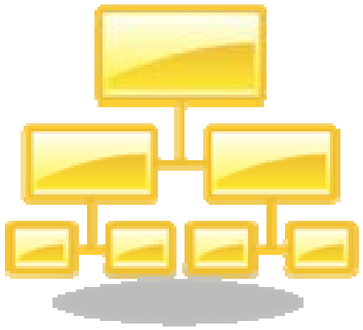
Statistical Approaches to Testing Ecological Models of Physical Activity Behavior

Ester Cerin, PhD
Baylor College of Medicine
Greg Norman, PhD
University of California, San Diego


Framework for determinants of physical activity and eating behavior
Secondary leverage points
Primary leverage points
Behavioral settings
Lifestyle
Enablers of choice
Social
Cultural
Psycho-biological Core
transportation system
food industry
architecture & building codes
political advocacy/ lobbying
exercise, physical activity & sports industry
recreation industry
health care industry
education system
entertainment industry
labor-saving device industry
information industry
government
local government
food stores
family
developers
property owners
restaurants and food outlets
recreation facilities
non-government organizations
nonprofit providers
community
shopping mall
health care providers
school board, districts
employer
food stores
health club
home
workplace
community activity providers
restaurants
religious, community and non-government organizations
parks, recreation centers, senior centers
vehicle of transport
shopping malls
neighborhood
day care
local school
social trends
seasonality
convenience
accessibility
situation or context – physical and social
source of information
cost
time
safety
knowledge
Lifestyle
social roles
habits
life stage
ethnic identities
interpersonal relationships
beliefs
values
educational attainment
life experience
socioeconomic status
self identities
pleasure
hierarchy of needs
genetics
physiology

Aims

- What models can we use with multilevel data?
- How do we interpret the results?
- What questions can we answer?
- What questions do you have for us?



PLACE (NQLS’ little brother)

GENERAL AIM OF PLACE …

To examine the relationships between physical attributes of the local community and transport-related and leisure time physical activity

- N = 2650 (aged 20-65) – Adelaide, Australia
- Multi-stage stratified sampling strategy
- 156 Census Collection Districts
  - High SES and high walkability
  - High SES and low walkability
  - Low SES and high walkability
  - Low SES and low walkability

Data Editor

Preserve | Restore | Sort | << | >> | Hide | Delete...

id[263] = D3768

	id	ccd	median_wk_~c	hhld_size	walkability2	sense_comm	informal_s~l	soc_cohesion	asingle	atown
263	D3768	4081001	5	2.3	20	10	17	11	Most	A
264	D4029	4081001	5	2.3	20	8	10	15	Most	N
265	D4031	4081001	5	2.3	20	10	19	16	Most	N
266	D4022	4081001	5	2.3	20	11	25	19	None	S
267	D3003	4081004	5	2.5	19	12	23	20	Some	S
268	D3937	4081004	5	2.5	19	9	19	16	Most	N
269	D2434	4081004	5	2.5	19	9	16	20	Most	A
270	D2601	4081004	5	2.5	19	6	5	7	Most	N
271	D3934	4081004	5	2.5	19	7	19	17	Most	N
272	D4016	4081004	5	2.5	19	12	20	12	Some	S
273	D3895	4081004	5	2.5	19	11	18	13	Most	N
274	D2441	4081004	5	2.5	19	10	18	17	Most	A
275	D3927	4081004	5	2.5	19	10	16	17	Most	A
276	D3943	4081004	5	2.5	19	6	18	12	Most	A
277	D3892	4081004	5	2.5	19	11	23	16	Most	N
278	D3978	4081004	5	2.5	19	8	16	17	All	N
279	D3965	4081004	5	2.5	19	10	15	13	All	N
280	D2061	4081004	5	2.5	19	11	13	16	Most	N
281	D2736	4081004	5	2.5	19	9	16	14	Most	A
282	D2444	4081004	5	2.5	19	[illegible]	[illegible]	[illegible]	[illegible]	N
283	D3907	4081004	5	2.5	19	[illegible]	[illegible]	[illegible]	[illegible]	A
284	D2579	40810[illegible]	[illegible]	[illegible]	19	10	18	17	Most	A
285	D2482	40810[illegible]	[illegible]	[illegible]	19	9	17	12	Most	N
286	D4073	4081102	5	2.7	21	10	18	18	Most	A
287	D2929	4081102	5	2.7	21	7	16	16	Most	N
288	D2093	4081102	5	2.7	21	10	19	15	Most	N
289	D4045	4081102	5	2.7	21	11	17	12	All	A
290	D4088	4081102	5	2.7	21	10	17	13	Most	N
291	D2837	4081102	5	2.7	21	11	22	19	All	N

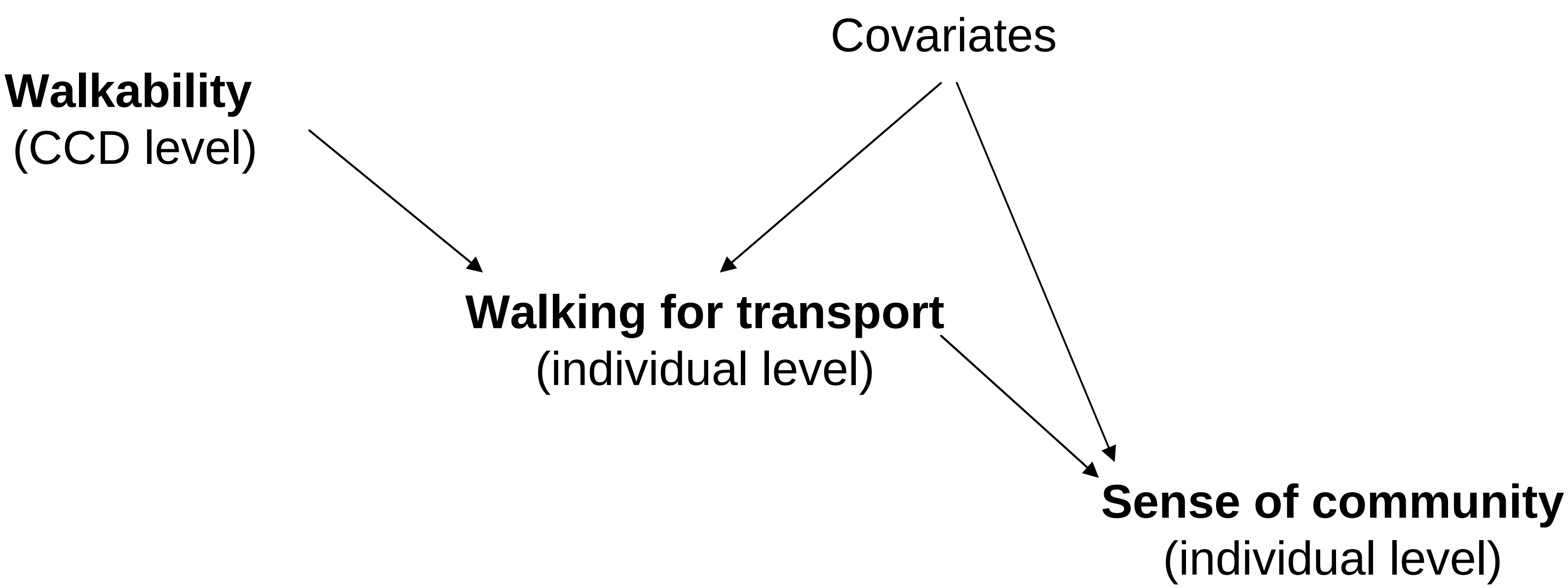
Area level variables

Respondent level variables

PLACE: variables

Census Collection District level	*Individual level*
GIS-based walkability	Weekly minutes of walking for transport
Median weekly income	Age
Median age	Gender
Average sense of community	Sense of community
??	Perceived access to services

Hypothetical model



Correlated data

- Clustering
- Intraclass correlation coefficient

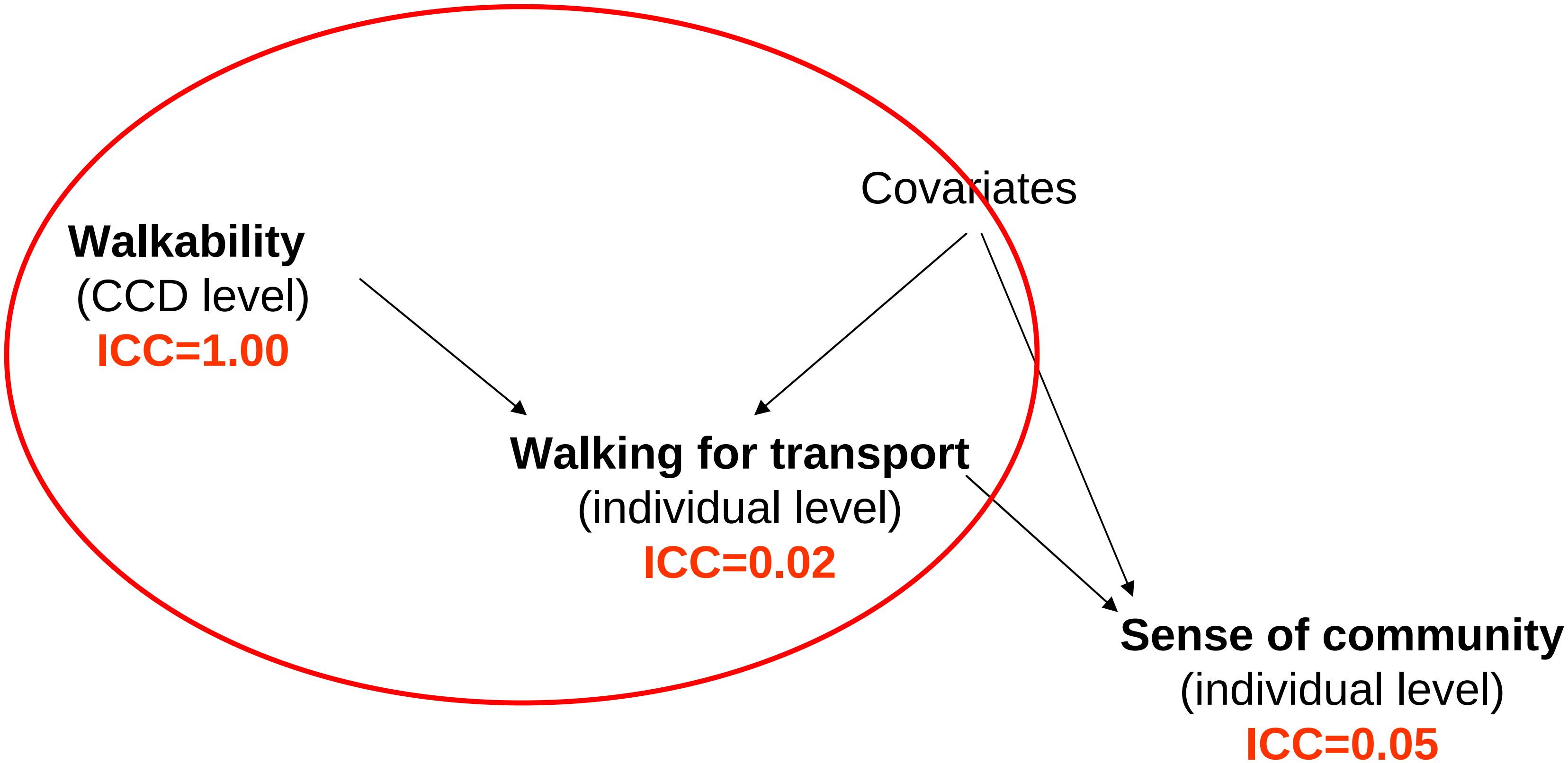
$$\rho = \frac{\sigma^2_{between-clusters}}{\sigma^2_{between-clusters} + \sigma^2_{within-clusters}}$$

- Observations are not independent
- Violation of independence assumption (independent and identically distributed errors) and use of standard statistical methods
  - Incorrect standard errors
  - Clustering primarily affects variance or precision of estimation rather than bias (unless individual-level associations between factors measured at the individual level differ from those at the area-level)

Effects of clustering (1)

- Comparisons between individuals grouped in clusters (e.g., area effect): less precise, or less informative, than comparisons made between the same number of completely independent individuals
- Individuals from the same cluster provide a smaller amount of information than completely independent individuals (random sample)
- The higher the ICC, the smaller the amount of information

Example: Effect of clustering (1)



```
. regress walktr1860 walkability2

      Source |       SS       df       MS              Number of obs =    2194
-------------+------------------------------           F(  1,  2192) =    8.39
       Model |  919780.993     1  919780.993           Prob > F      =  0.0038
    Residual |   240400149  2192  109671.601           R-squared     =  0.0038
-------------+------------------------------           Adj R-squared =  0.0034
       Total |   241319930  2193  110041.008           Root MSE      =  331.17

------------------------------------------------------------------------------
  walktr1860 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
walkability2 |   2.634455   .9096947     2.90   0.004     .8505011    4.418409
       _cons |   159.7848   16.65254     9.60   0.000     127.1284    192.4412
------------------------------------------------------------------------------
```

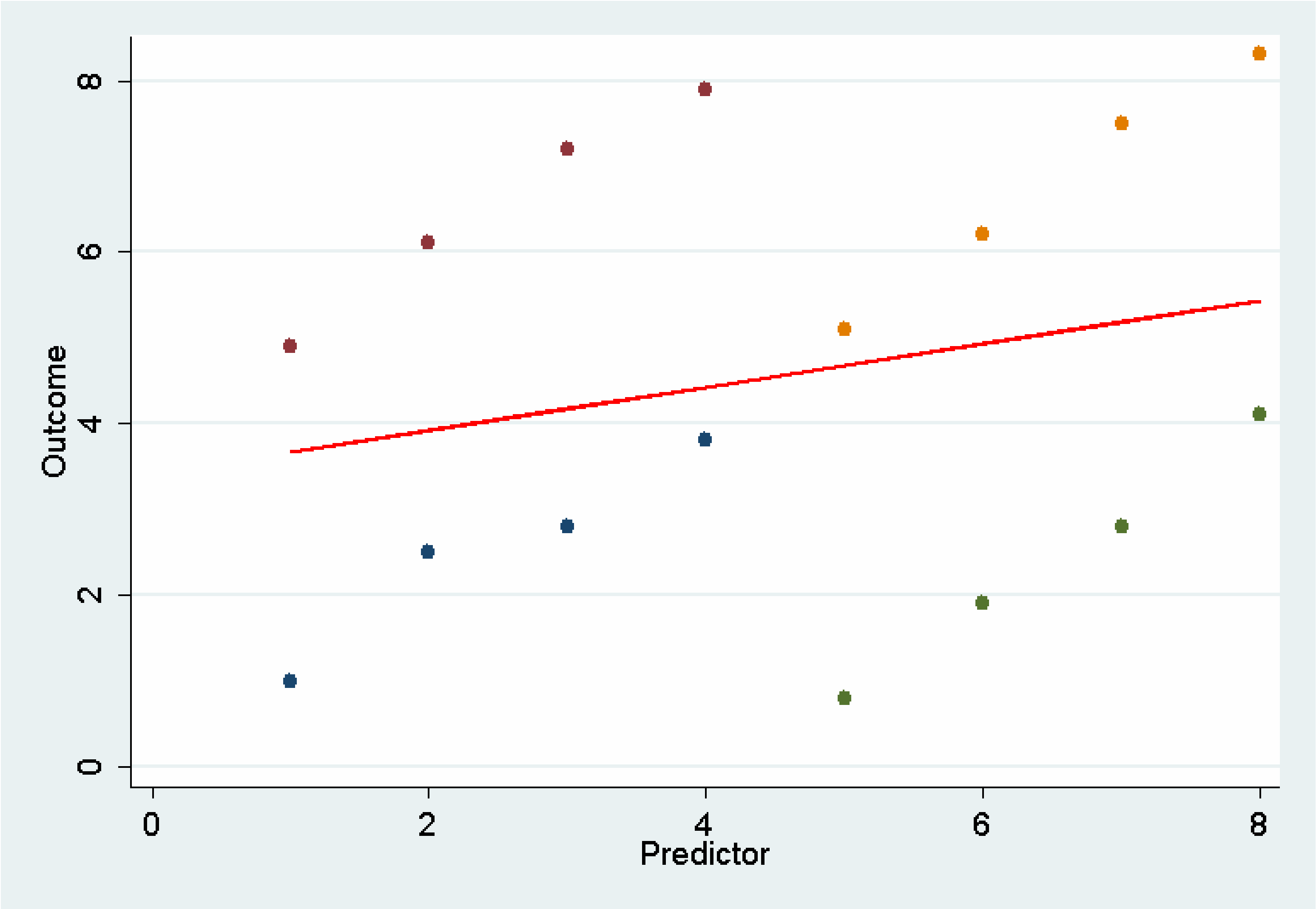
```
Linear regression                                      Number of obs =    2194
                                                       F(  1,   153) =    6.68
                                                       Prob > F      =  0.0107
                                                       R-squared     =  0.0038
Number of clusters (ccd) = 154                         Root MSE      =  331.17

------------------------------------------------------------------------------
             |               Robust
  walktr1860 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
walkability2 |   2.634455   1.019193     2.58   0.011     .6209474    4.647963
       _cons |   159.7848   16.79041     9.52   0.000     126.6138    192.9558
------------------------------------------------------------------------------
```

Effects of clustering (2)

- Comparisons between individuals within a cluster: more precise, or more informative, than comparisons made between individuals in different clusters
- If the variation between observations is less within clusters than between, then by just comparing within those clusters we should be able to see differences more clearly

Within area effects

ICC= 0.75



```
. regress y x

      Source |       SS       df       MS              Number of obs =      16
-------------+------------------------------           F(  1,    14) =    0.88
       Model |  5.37574409     1  5.37574409           Prob > F      =  0.3653
    Residual |  85.9636315    14   6.1402594           R-squared     =  0.0589
-------------+------------------------------           Adj R-squared = -0.0084
       Total |  91.3393756    15  6.08929171           Root MSE      =   2.478

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .2529762    .270367     0.94   0.365    -.3269034    .8328558
       _cons |   3.417857   1.365287     2.50   0.025     .4896087    6.346105
------------------------------------------------------------------------------
```

```
                              (Std. Err. adjusted for clustering on ccd)
------------------------------------------------------------------------------
             |             Semi-robust
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.008214   .0496541    20.30   0.000      .910894    1.105534
       _cons |   .0192862   1.740609     0.01   0.991    -3.392244    3.430817
------------------------------------------------------------------------------
```

Modeling multilevel data:
‘Ordinary’ single-level regression

Inclusion of dummy variables representing clusters + their interaction with the predictor of interest

```
. reg y x  dummy1 dummy2 dummy3 xd1 xd2 xd3

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |       1.09   .0981071    11.11   0.000     .8637648    1.316235
      dummy1 |   .6600007   .7006246     0.94   0.374    -.9556425    2.275644
      dummy2 |   4.310001   .7006246     6.15   0.000     2.694357    5.925644
      dummy3 |  -4.309999   .9150818    -4.71   0.002    -6.420182   -2.199817
         xd1 |  -.2200001   .1387443    -1.59   0.151    -.5399452    .0999449
         xd2 |  -.0800001   .1387443    -0.58   0.580    -.3999451    .2399449
         xd3 |  -.0100001   .1387443    -0.07   0.944    -.3299452    .3099449
       _cons |  -.3100007   .6470605    -0.48   0.645    -1.802125    1.182124
------------------------------------------------------------------------------
```

```
. reg y x  dummy1 dummy2 dummy3

      Source |       SS       df       MS              Number of obs =      16
-------------+------------------------------           F(  4,    11) =  462.94
       Model |  90.8000007     4  22.7000002           Prob > F      =  0.0000
    Residual |  .539374949    11  .049034086           R-squared     =  0.9941
-------------+------------------------------           Adj R-squared =  0.9919
       Total |  91.33[illegible]6    15  6.08929171           Root MSE      =  .22144

------------------------------------------------------------------------------
           y |  [illegible]   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |     1.0125   .0495147    20.45   0.000     .9035189    1.121481
      dummy1 |        -.2   .2524764    -0.79   0.445    -.7556967    .3556968
      dummy2 |        3.8   .2524764    15.05   0.000     3.244303    4.355697
      dummy3 |     -4.375   .1565792   -27.94   0.000    -4.719628   -4.030372
       _cons |   .1937499   .3403572     0.57   0.581    -.5553713    .9428711
------------------------------------------------------------------------------
```

Modeling multilevel data:
'Ordinary' single-level regression

- Advantages
  - Easy
  - Does not require specialized statistical software
- Disadvantages
  - Loss of power and efficiency
    - With 154 CCDs and 1 predictor of interest we would require 306 variables!!
  - Cannot simultaneously estimate predictors' effects and outcome variance attributable to different levels of variation (area and individual)

Modeling multilevel data:
Single-level regression with robust standard errors (sandwich; Huber-White estimators)

- **Advantages:** easy to use
- OK to use when:
  - Examining area effects (predictors are measured at the area level / or aggregated at the area level)
  - Examining associations between individual-level variables AND area- & individual-level effects are similar, OR both contextual and individual level effects are included in the model
  - Requires a large number of clusters (N > 20!)
- **Disadvantages:**
  - Inefficient
  - Cannot simultaneously estimate predictors' effects and outcome variance attributable to different levels of variation (area and individual)
  - Cannot account for more than 2 levels of variation

Contextual effects???

A predictor’s average value for each cluster

ID	Area	Y	X	Mean(X)	X - Mean(X)
1	1	12	3	4	-1
2	1	13	5	4	1
3	2	11	4	3.5	0.5
4	2	10	3	3.5	-0.5
5	3	13	6	7	-1
6	3	16	8	7	1

A bit of scary math …

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 \bar{X}_j$$

Individual-level effect

Contextual effect

The 'mystery' model

(Std. Err. adjusted for clustering on ccd)

y	Coef.	Semi-robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
x	1.008214	.0496541	20.30	0.000	.910894	1.105534
_cons	.0192862	1.740609	0.01	0.991	-3.392244	3.430817

Single-level regression with robust standard errors

y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
x	.2529762	.47291	0.53	0.630	-1.252035	1.757987
_cons	3.417857	2.380319	1.44	0.247	-4.157381	10.9931

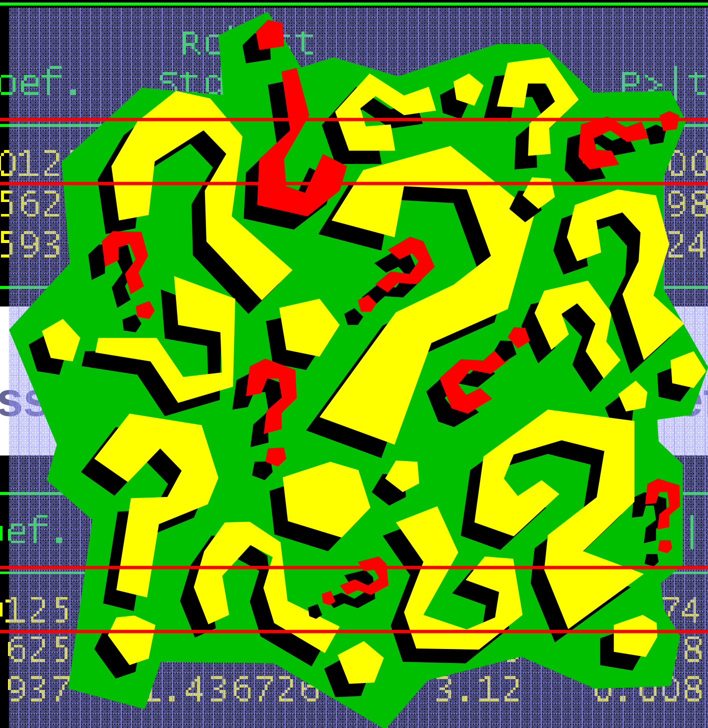
Single-level regression with robust standard errors, adjusted for contextual effects

y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
x_diff	1.0125	.0544863	18.58	0.000	.8391004	1.1859
mean_x	.015625	.6498948	0.02	0.982	-2.05263	2.08388
_cons	4.485937	3.092385	1.45	0.243	-5.355413	14.32729

What if we do not account for clustering effects (do not use robust standard errors)?

Single-level regression with robust standard errors, adjusted for contextual effects

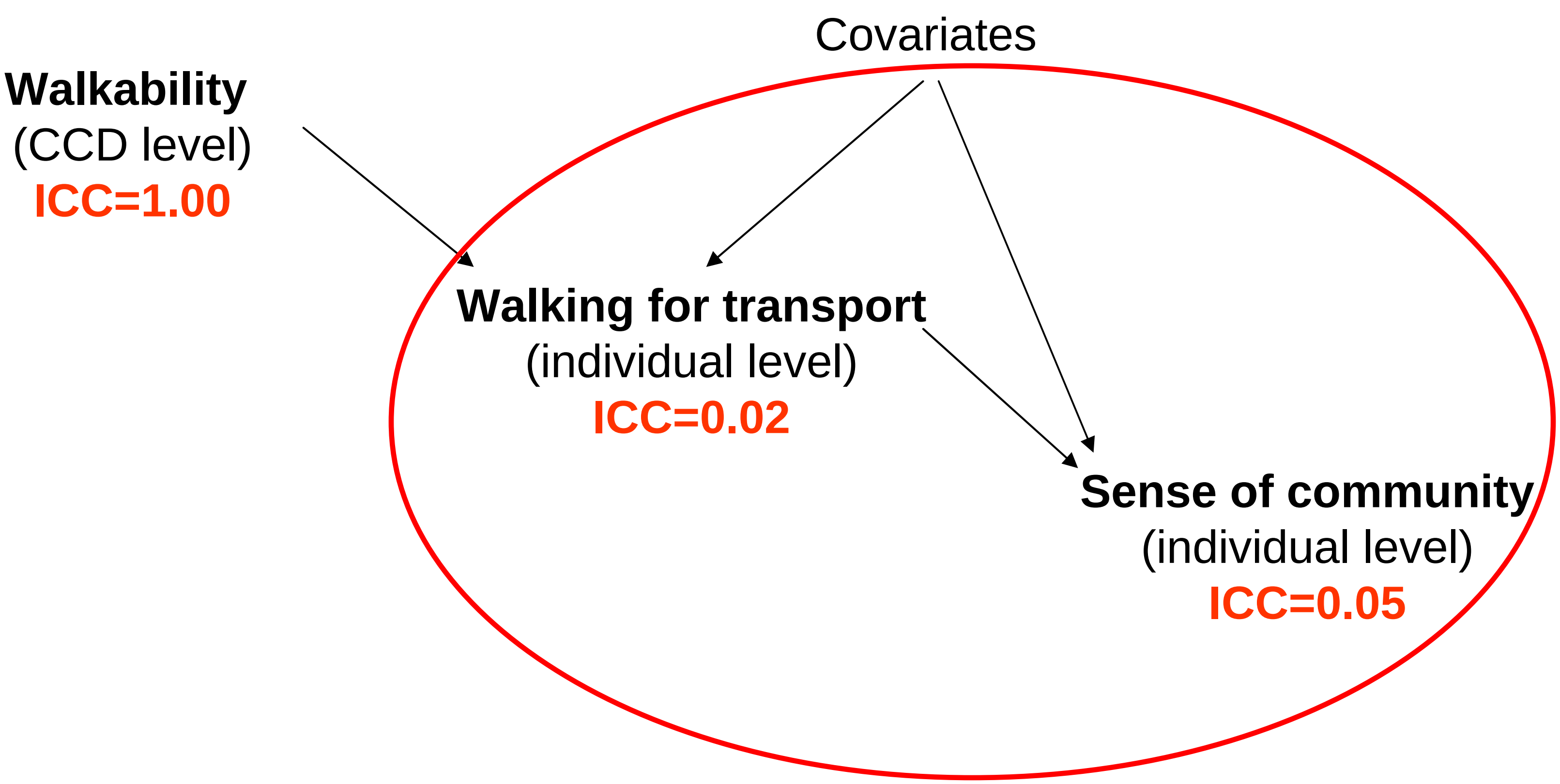
y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
x_diff	1.012	[illegible]	[illegible]	000	.8391004	1.1859
mean_x	.01562	[illegible]	[illegible]	982	-2.05263	2.08388
_cons	4.48593	[illegible]	[illegible]	243	-5.355413	14.32729



Single-level regress [illegible] ffects

y	Coef.	[illegible]	[illegible]	[illegible]	[95% Conf.	Interval]
x_diff	1.0125	[illegible]	[illegible]	74	-.1150117	2.140012
mean_x	.015625	[illegible]	[illegible]	8	-.6146732	.6459232
_cons	4.485937	1.436726	3.12	0.008	1.382079	7.589796

Real data:



Sense of community = Intercept + Walking

The ‘mystery’ model

(Std. Err. adjusted for clustering on ccd)

sense_comm	Coef.	Semi-robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
walktr1860	.0003305	.0001477	2.24	0.025	.000041	.0006199
_cons	9.93914	.0659446	150.72	0.000	9.809891	10.06839

Single-level regression with robust standard errors

sense_comm	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
walktr1860	.0003552	.0001484	2.39	0.018	.000062	.0006484
_cons	9.946654	.0649563	153.13	0.000	9.818327	10.07498

Single-level regression with robust standard errors, adjusted for contextual effects

sense_comm	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
walk_d	.0002901	.000153	1.90	0.060	-.0000121	.0005922
walk_mean	.0010753	.0005829	1.84	0.067	-.0000761	.0022268
_cons	9.800504	.126766	77.31	0.000	9.550066	10.05094

How can we get sandwich SE?

- Stata
  - Use option *robust* or *cluster(…)*
- SAS
  - proc GENMOD; statement REPEATED
- SPSS
  - Generalized Linear Models -> Covariance Matrix -> Robust estimator
- R or S-Plus
  - geeglm (geepack) -> std.err="san.se"

The ‘mystery’ model

Generalized Estimating Equations
(marginal model approach)

GEE

- Work hard at the correct modeling of the mean structure (predictors) while using methods of estimation that are valid in the presence of correlation and robust to potential misspecification of the detail of the covariance structure.
- Marginal modeling = focus on the model of the marginal mean of *y*, with the covariance structure treated as a nuisance
- Weighs clustered data and makes estimation more efficient

Scary math again …

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

Independent observations

Independence working correlation matrix (equivalent to 'common' regression)

$$\sigma^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Yep … one more fright …

Correlated observations

$$\sigma^2 \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

Exchangeable WCM
(no logical ordering for observations within a cluster)

$$\sigma^2 \begin{pmatrix} 1 & \rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \rho & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \rho & \rho & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho & 1 & \rho & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho & \rho & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \rho & \rho \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \rho & 1 & \rho \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \rho & \rho & 1 \end{pmatrix}$$

Other WCM:
-Unstructured
-Auto-regressive
-Fixed
-M-dependent

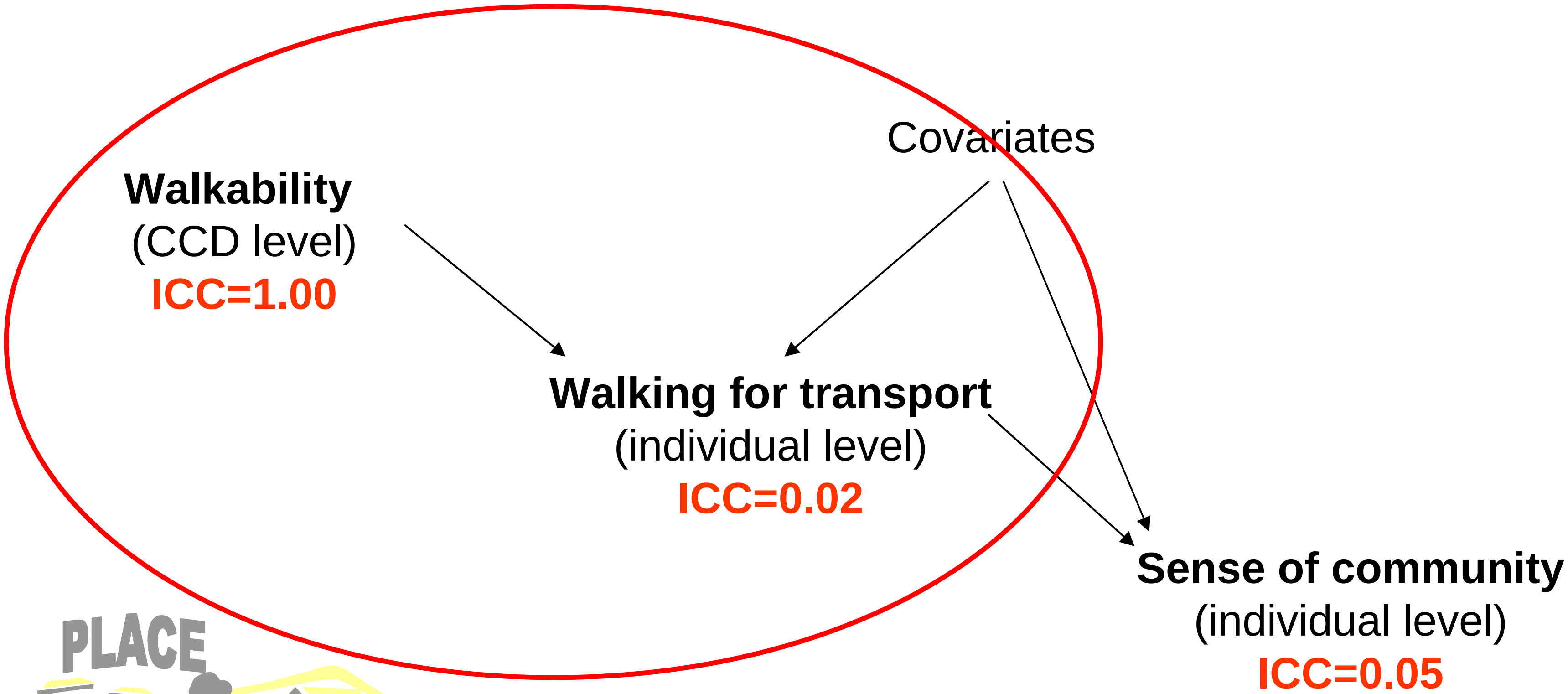
Modeling multilevel data:

Generalized estimating equations (with and without sandwich; Huber-White estimators)

- Advantages
  - Relatively easy to use
  - Give marginal estimates of effects: gives estimates that correspond to comparing two observations randomly selected from the population (without matching on clusters)
  - Public health significance
- Disadvantages
  - Widely unbalanced clusters create problems
  - Cannot simultaneously estimate predictors' effects and outcome variance attributable to different levels of variation (area and individual)
  - Cannot get estimates of variability in effects (standard deviations of slopes)
  - Cannot be applied to data with more than 2 levels of variation

Real data:

Generalized estimating equations (with and without sandwich; Huber-White estimators)



Physical Activity in Localities and Community Environments

Real data:

Generalized estimating equations (with and without sandwich; Huber-White estimators)

- Single-level regression with dummy variables? Why not?
- Single-level regression with sandwich estimators? Why not?
- Do we need to model individual and contextual effects of walkability?
- What working correlation matrix shall we use?
- Can we use sandwich estimators of SE?
- What about the distribution of the outcome variable? Is it skewed or normally distributed? (GLM handout)
- What about the shape of the relationship between the predictor and the outcome? (GLM handout)

PLACE

Physical Activity in Localities and Community Environments

Real data:

Generalized estimating equations (with and without sandwich; Huber-White estimators)

Single-level with robust SE

(Std. Err. adjusted for 154 clusters in ccd)

walktr1860	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
walkability2	2.708121	1.035585	2.62	0.009	.6784127	4.73783
aa_age	1.248108	.4084961	3.06	0.002	.4474699	2.048745
_cons	101.628	26.14484	3.89	0.000	50.38501	152.8709

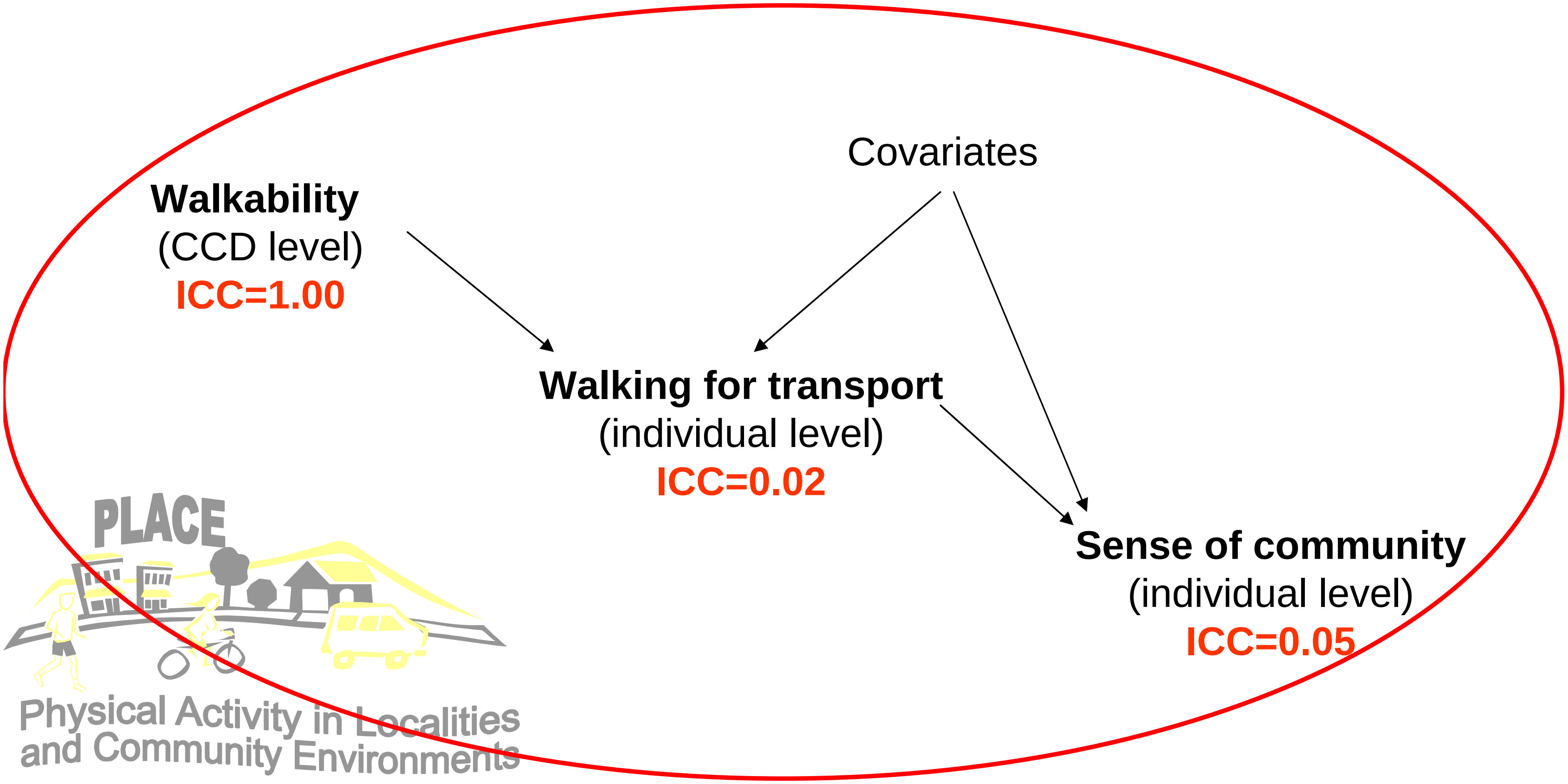
GEE with robust SE

(Std. Err. adjusted for clustering on ccd)

walktr1860	Coef.	Semi-robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
walkability2	2.774389	.9787022	2.83	0.005	.8561676	4.69261
aa_age	1.25721	.517342	2.43	0.015	.2432384	2.271182
_cons	100.6298	30.76888	3.27	0.001	40.32388	160.9357

Real data:

Generalized estimating equations (with and without sandwich; Huber-White estimators)



Real data:

Generalized estimating equations (with and without sandwich; Huber-White estimators)

- Do we need to model individual and contextual effects of walking?
- What working correlation matrix shall we use?
- What about the distribution of the outcome variable? Is it skewed or normally distributed?
- What about the shape of the relationship?

PLACE
Physical Activity in Localities and Community Environments

Real data:

Generalized estimating equations (with and without sandwich; Huber-White estimators)

Single-level with robust SE

(Std. Err. adjusted for 154 clusters in ccd)

sense_comm	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
walktr1860	.0002724	.0001483	1.84	0.066	-.0000183	.0005631
walkability2	.0179439	.0069821	2.57	0.010	.0042593	.0316285
aa_age	.0295024	.0033922	8.70	0.000	.0228538	.0361511
_cons	8.324257	.2077744	40.06	0.000	7.917026	8.731487

GEE with robust SE

(Std. Err. adjusted for clustering on ccd)

sense_comm	Coef.	Semi-robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
walktr1860	.0002595	.0001485	1.75	0.080	-.0000315	.0005506
walkability2	.0188467	.0069958	2.69	0.007	.0051352	.0325581
aa_age	.0288185	.0034346	8.39	0.000	.0220868	.0355502
_cons	8.332918	.2110034	39.49	0.000	7.919359	8.746477

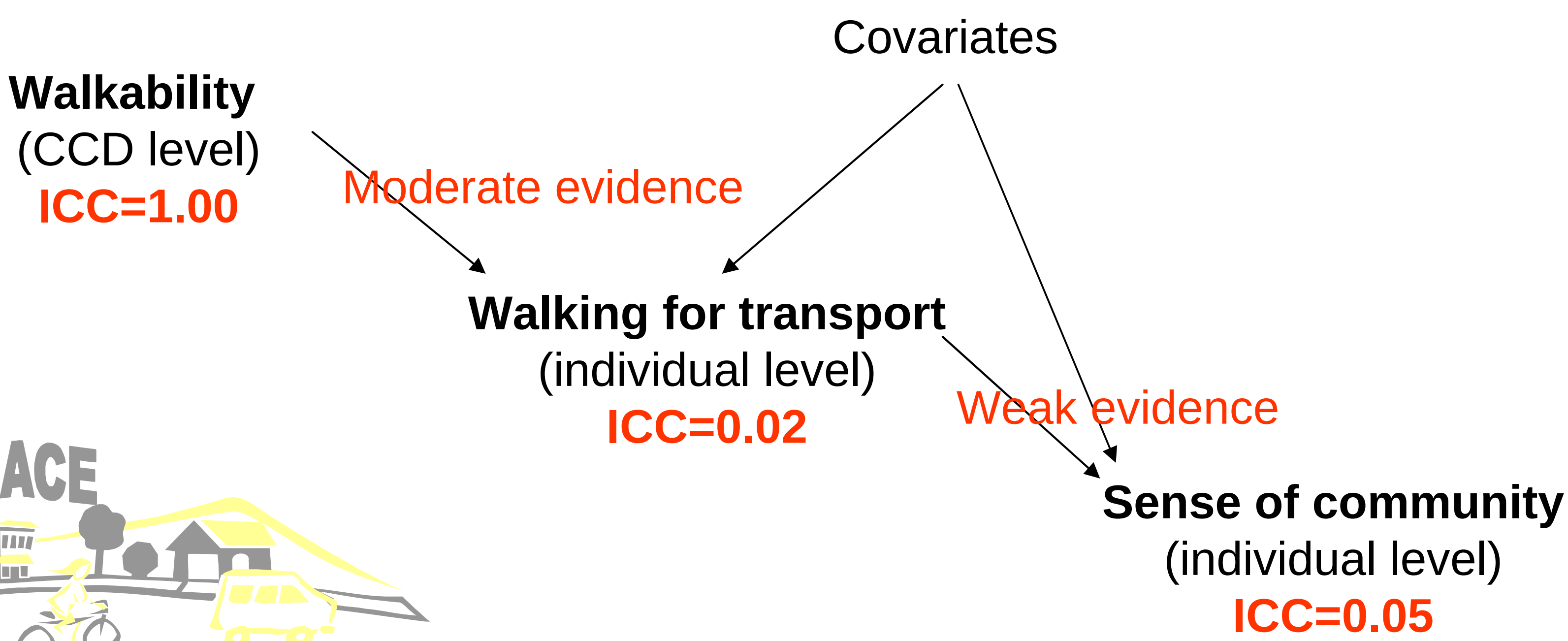
Does walking for transport mediate the relationship between walkability and sense of community?

Weak evidence …

PLACE
Physical Activity in Localities and Community Environments

Real data:

Generalized estimating equations (with and without sandwich; Huber-White estimators)



Physical Activity in Localities and Community Environments

Software for GEE

- SPSS
  - Generalized Linear Models -> GEE
- Stata
  - xtgee
- SAS
  - proc GENMOD; statement REPEATED
- R or S-Plus
  - gee (Vincent Carey)

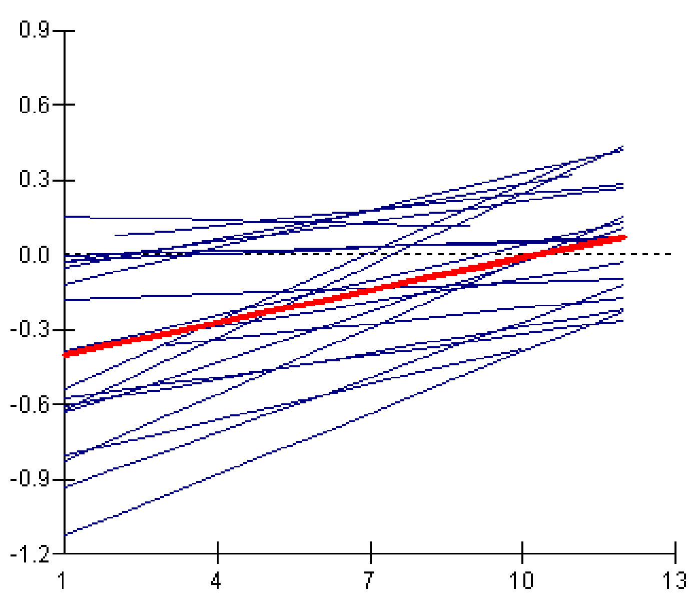
Main components of GEE syntax

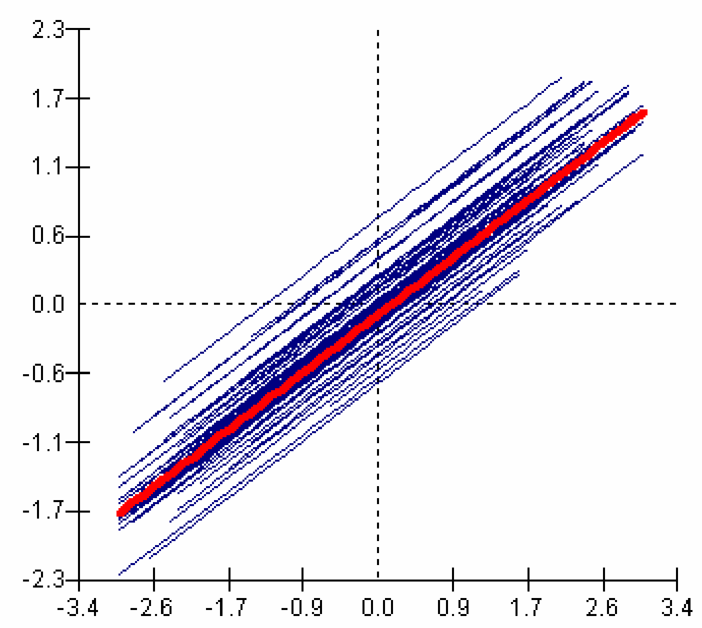
- Outcome and explanatory variables
- Variance function (normal; gamma; binomial, negative binomial …)
- Link function (identity, logarithmic, logit, inverse, power, probit …)
- Specify the cluster variable
- Identify the working correlation matrix
- Model-based or empirical (robust) standard errors

```
xtgee walktr1860 walkability2 aa_age, i(ccd) f(gamma) link(identity) corr(exch) robust
```

GEEE … and what about multilevel linear models?

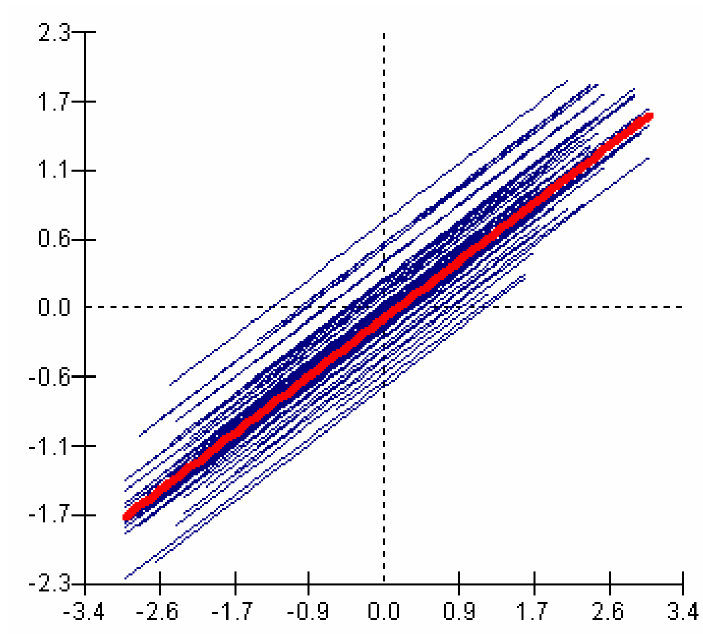
- … also called hierarchical linear models
- … OR linear mixed models
- … OR generalized linear mixed models





… MLM / GLMM / HLM … whatever …

Random intercept model



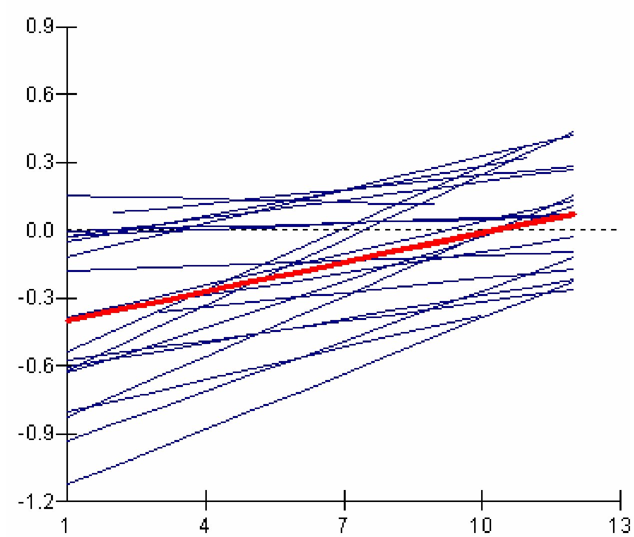
$$Walking_{ij} = \beta_{0j}cons + \beta_1(age)_{ij} + \beta_2(walkability)_j$$

$$\beta_{0j} = \beta_0 + u_{0j} + e_{0ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \quad e_{0ij} \sim N(0, \sigma_{e0}^2)$$

… MLM / GLMM / HLM … whatever …

Random intercept and random slope model



$$Sense\_Community_{ij} = \beta_{0ij} cons + \beta_1 (age)_{ij} + \beta_2 (walkability)_j + \beta_{3j} (walking)_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j} + e_{0ij}$$

$$\beta_{3j} = \beta_3 + u_{1j}$$

$$u_{0j} \sim N(0, \sigma^2_{u0}) \quad e_{0ij} \sim N(0, \sigma^2_{e0})$$

$$u_{1j} \sim N(0, \sigma^2_{u1}) \quad Cov(u_{0j}, u_{1j}) = \rho \sigma_{u0} \sigma_{u1}$$

GEEE … and what about multilevel linear models?

- Advantages over GEE
  - More robust in case of missing data
  - More robust in case of unbalanced clusters
  - Can estimate variances at different levels (individual versus area)
    - How much do the effects of walking for transport on sense of community vary across CCDs?
    - What explains such variations?
- Disadvantages
  - Conditional, area-specific effects
  - More difficult to set up
  - Specialized software (especially for more than 3 levels of variation)

… so what about multilevel linear models?

```
log likelihood = -15838.343

------------------------------------------------------------------------------
   walktr1860 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
 walkability2 |   2.784656   .9782341     2.85   0.004     .8673528     4.70196
       aa_age |   1.525122   .5994114     2.54   0.011     .3502977    2.699947
_Iaa_gende_1  |  -11.91044   14.90564    -0.80   0.424    -41.12495    17.30407
        _cons |   96.04954   35.66595     2.69   0.007     26.14556    165.9535
------------------------------------------------------------------------------

Variance at level 1
------------------------------------------------------------------------------
  108169.12 (3359.1819)

Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (ccd)

    var(1): 999.16976 (909.73504)
------------------------------------------------------------------------------
```

… so what about multilevel linear models?

```
log likelihood = -4560.5175

------------------------------------------------------------------------------
  sense_comm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  walktr1860 |    .000258   .0001291     2.00   0.046     4.93e-06     .000511
walkability2 |   .0189959   .0070668     2.69   0.007     .0051452    .0328467
      aa_age |   .0287477   .0036222     7.94   0.000     .0216483     .035847
_Iaa_gende_1 |  -.0024575   .0894361    -0.03   0.978     -.177749    .1728341
       _cons |   8.334155   .2263955    36.81   0.000     7.890428    8.777882
------------------------------------------------------------------------------

Variance at level 1
------------------------------------------------------------------------------
  3.7862908 (.1190439)

Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (ccd)

    var(1): .16177849 (.05232494)
------------------------------------------------------------------------------
```

Does walking for transport mediate the relationship between walkability and sense of community?

Some evidence …

PLACE
Physical Activity in Localities and Community Environments

Does walking for transport mediate the relationship between walkability and sense of community in all CCDs?

PLACE

Physical Activity in Localities and Community Environments

… so what about multilevel linear models?

```
log likelihood = -4557.9158

------------------------------------------------------------------------------
  sense_comm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  walktr1860 |   .0002584   .0001465     1.76   0.078    -.0000288    .0005456
walkability2 |   .0205094   .0071913     2.85   0.004     .0064148    .0346041
      aa_age |   .0290625   .0036118     8.05   0.000     .0219835    .0361414
_Iaa_gende_1 |  -.0029531   .0893543    -0.03   0.974    -.1780843    .1721781
       _cons |   8.293181   .2289085    36.23   0.000     7.844528    8.741833
------------------------------------------------------------------------------

Variance at level 1
------------------------------------------------------------------------------
  3.7317409 (.11978565)

Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (ccd)

    var(1): .23560551 (.07392603)
    cov(1,2): -.00021753 (.00012073) cor(1,2): -.64751056

    var(2): 4.790e-07 (2.961e-07)
------------------------------------------------------------------------------
```

CCD-level variation in slope of walking for transport: 0.000258±0.000692

What explains variations in effects of walking for transport on sense of community across CCDs?

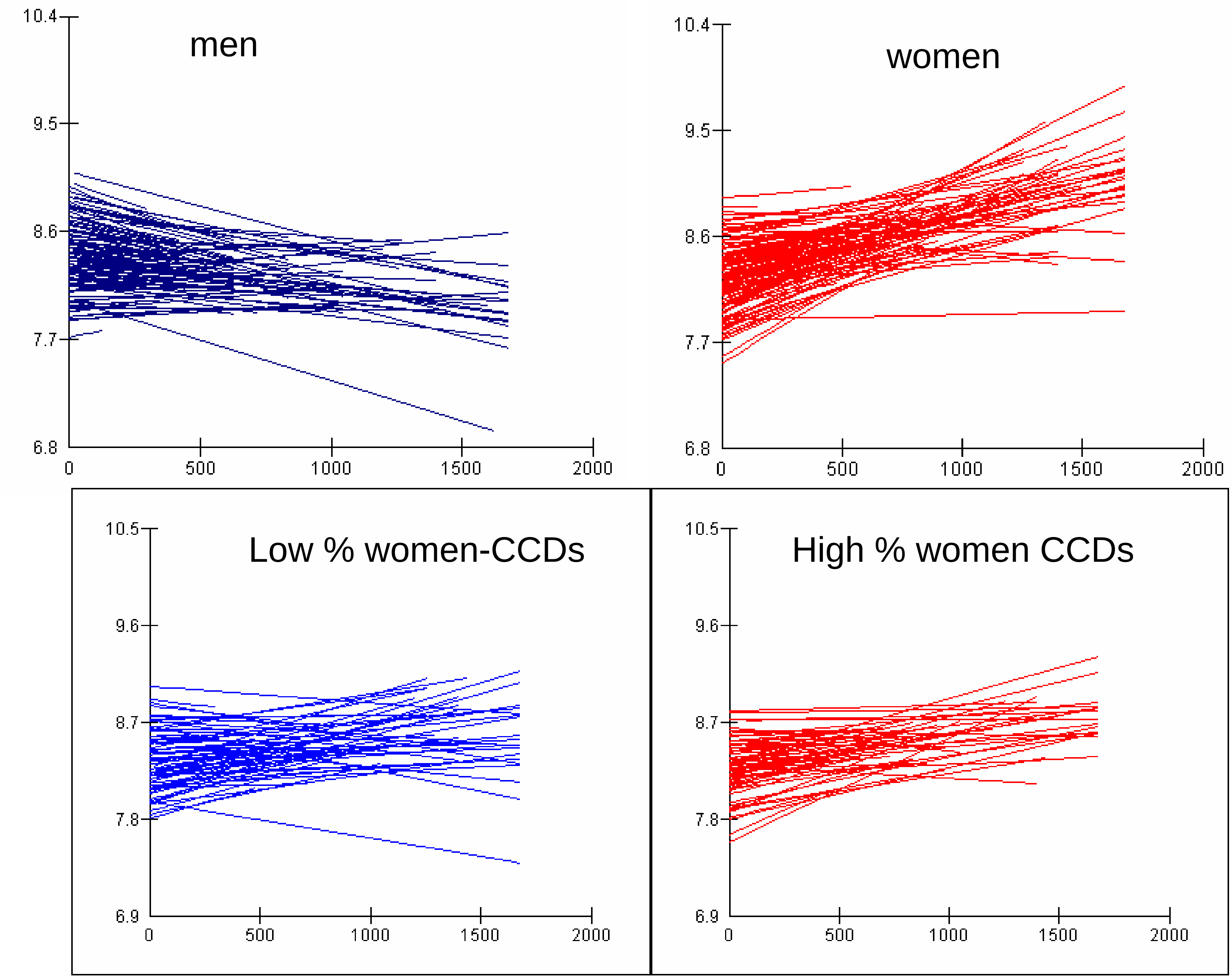
PLACE

Physical Activity in Localities and Community Environments

… so what about multilevel linear models?

sense_comm	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_Iaa_gende_1	-.1492161	.1044089	-1.43	0.153	-.3538538	.0554217
walktr1860	-.0001944	.0002221	-0.88	0.381	-.0006297	.0002409
_Iaa_Xwalk~1	.0007364	.0002727	2.70	0.007	.0002018	.0012709
walkability2	.0202918	.0071281	2.85	0.004	.0063209	.0342626
aa_age	.0293789	.0036078	8.14	0.000	.0223077	.03645
_cons	8.375554	.2300588	36.41	0.000	7.924647	8.82646

```
Variance at level 1
------------------------------------------------------------------------------
  3.7229165 (.11952071)

Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (ccd)

    var(1): .23594917 (.07346059)
    cov(1,2): -.00022435 (.00011921) cor(1,2): -.68126111

    var(2): 4.596e-07 (2.917e-07)
------------------------------------------------------------------------------
```


men
women
Low % women-CCDs
High % women CCDs
10.4
9.5
8.6
7.7
6.8
10.5
9.6
8.7
7.8
6.9
0
500
1000
1500
2000

PLACE
Physical Activity in Localities and Community Environments

Real data:
MLM / GLMM / HLM … whatever …

Walkability
(CCD level)

2.78*

Walking for transport
(individual level)

Sufficient evidence
… but very small effect…

0.0005*

-0.0002

Sense of community
(individual level)

Insufficient evidence

Software for MLM

- SPSS
  - (Linear) Mixed Models
- Stata
  - xtreg; xtlogit; xtETC; gllamm (Generalized Linear Latent and Mixed Models)
- SAS
  - proc MIXED
- R or S-Plus
  - nlme (non-linear mixed effects)
- MLwiN (University of Bristol, UK)

Conclusions

- When shall we use single-level regression with dummy variables representing clusters?
- When shall we use GEE?
- When can we simply use sandwich estimators of SE?
- When would we prefer the multilevel linear models?