



# **Harmonizing Measures Using Individual Participant Data**

**George W. Howe  
George Washington University**



## In collaboration with

- C. Hendricks Brown
- Getachew Dagne
- Tatiana Perrino
- Hilda Pantin
- Ahnalee Brincks
- Gracelyn Cruden
- Juned Siddique
- Shi Huang
- Robert Gibbons
- William Beardslee
- Irwin Sandler



## Growth in data mining efforts

- Combining data from multiple trials, in order to:
  - Test stability of effects over trials, times, populations, investigators
  - Detect effects requiring large samples (eg. Moderator effects)
- Methods include
  - Meta-analysis with summary statistics
  - Meta-analysis with individual participant data (IDA: integrated data analysis: Curran & Husong, 2009)



## Key Measurement Assumptions

- Measure invariance
  - When all trials include the same measure
- Assumes:
  - Measure is assessing the same thing in all trials
  - The metric or scale is the same across all trials (when testing for moderation)

## Testing Measurement Invariance

- Some tests can be done on summary statistics
  - Eg. Meta-analysis of reliability coefficients
- More rigorous tests require individual participant data
  - Do indicators operate in same way across all studies?
  - Moderated nonlinear factor analysis (Bauer & Hussong, 2009)
  - Recent extensions by Henderson, Greenbaum, and Wei



## Construct Equivalence

- Single measure invariance is subset of the broader assumption of construct equivalence
- Important when different studies use different measures of the same construct

## Example

Depression measures in prevention trials contributing to data synthesis project

|                            | CDI | CDRS | CESD | CBCL | YSR |
|----------------------------|-----|------|------|------|-----|
| <b>Bridges</b>             |     |      |      | Yes  | Yes |
| <b>Familias Unidas II</b>  |     |      |      |      | Yes |
| <b>Familias Unidas DJJ</b> |     |      |      |      | Yes |
| <b>Family Bereavement</b>  | Yes |      |      | Yes  |     |
| <b>Family Talk</b>         | Yes |      |      |      | Yes |
| <b>K-IPT</b>               |     | Yes  | Yes  |      |     |
| <b>Narstad-IPT</b>         |     | Yes  | Yes  |      |     |
| <b>New Beginnings</b>      | Yes |      |      | Yes  |     |
| <b>Project Alliance 1</b>  | Yes |      |      | Yes  | Yes |
| <b>Project Alliance 2</b>  | Yes |      |      |      |     |

# Construct Equivalence

- Assumes:
  - All measures index the same construct in the same way across all studies.
  - Effect sizes are in the same metric across all measures across all studies
- Current practices for dealing with construct equivalence in meta-analysis with varying outcome measures:
  - Just assume it holds: based on expert opinion, received wisdom, or even same label
  - Occasionally use meta-regression to study whether effects differ by measure
- How can we be more rigorous?





## Harmonizing: methods for establishing construct equivalence

- For mining existing data
  - Find common index with common metrics, based on existing data from one or more measures across multiple studies
- For establishing common measurement approaches for future studies
  - Use current data to identify stronger measures for future research
  - Develop common measurement approaches that will facilitate future synthesis



## Harmonizing construct measurement in existing data

- Consider problem from missing data or potential outcomes framework
  - Scale level: Use available information to impute scores (Juned Siddique and Ahnalee Brincks are developing these methods)
  - Item level: Use full information methods with multilevel IRT models



## Data Structure for Pool of Trials

- Multilevel problem:
  - Items nested within measures
  - Nested within individuals
  - (Nested within families)
  - Nested within intervention conditions
  - Nested within trials

# Data Structure

|          |         |          |          | Construct 1 |       |       |       |           |       |       |       |           |       |       |       |  |
|----------|---------|----------|----------|-------------|-------|-------|-------|-----------|-------|-------|-------|-----------|-------|-------|-------|--|
|          |         |          |          | Measure 1   |       |       |       | Measure 2 |       |       |       | Measure 3 |       |       |       |  |
|          |         |          |          | ITM11       | ITM12 | ITM13 | ITM14 | ITM21     | ITM22 | ITM23 | ITM24 | ITM31     | ITM32 | ITM33 | ITM34 |  |
| Trial 1  | EXP Grp | Family 1 | Member 1 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 2 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 3 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         | Family 2 | Member 1 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 2 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 3 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          | CTL Grp | Family 1 | Member 1 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 2 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 3 |             |       |       |       |           |       |       |       |           |       |       |       |  |
| Family 2 |         | Member 1 |          |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         | Member 2 |          |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         | Member 3 |          |             |       |       |       |           |       |       |       |           |       |       |       |  |
| Trial 2  | EXP Grp | Family 1 | Member 1 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 2 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 3 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         | Family 2 | Member 1 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 2 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 3 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          | CTL Grp | Family 1 | Member 1 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 2 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 3 |             |       |       |       |           |       |       |       |           |       |       |       |  |
| Family 2 |         | Member 1 |          |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         | Member 2 |          |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         | Member 3 |          |             |       |       |       |           |       |       |       |           |       |       |       |  |
| Trial 3  | EXP Grp | Family 1 | Member 1 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 2 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 3 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         | Family 2 | Member 1 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 2 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 3 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          | CTL Grp | Family 1 | Member 1 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 2 |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         |          | Member 3 |             |       |       |       |           |       |       |       |           |       |       |       |  |
| Family 2 |         | Member 1 |          |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         | Member 2 |          |             |       |       |       |           |       |       |       |           |       |       |       |  |
|          |         | Member 3 |          |             |       |       |       |           |       |       |       |           |       |       |       |  |

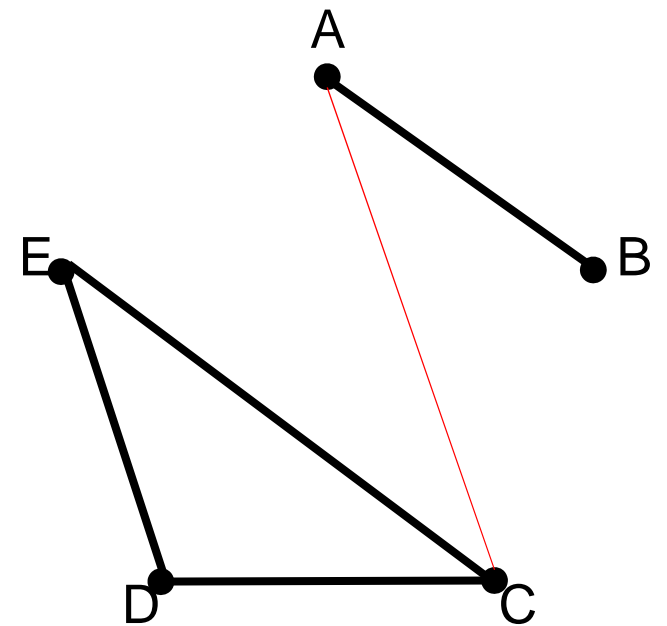
## Example: Collaborative Data Synthesis Project (CDS)

- Prevention trials to reduce risk of adolescent problems
  - Targets varied: family interaction, adolescent coping
  - All included measures of depressive symptoms
  - 12 trials, total of 3173 adolescents
  - 5 different measures
    - CDI
    - CDRS
    - CESD
    - YSR
    - CBCL
  - Here we focus on pretest measurement of depression

# CDS Data:

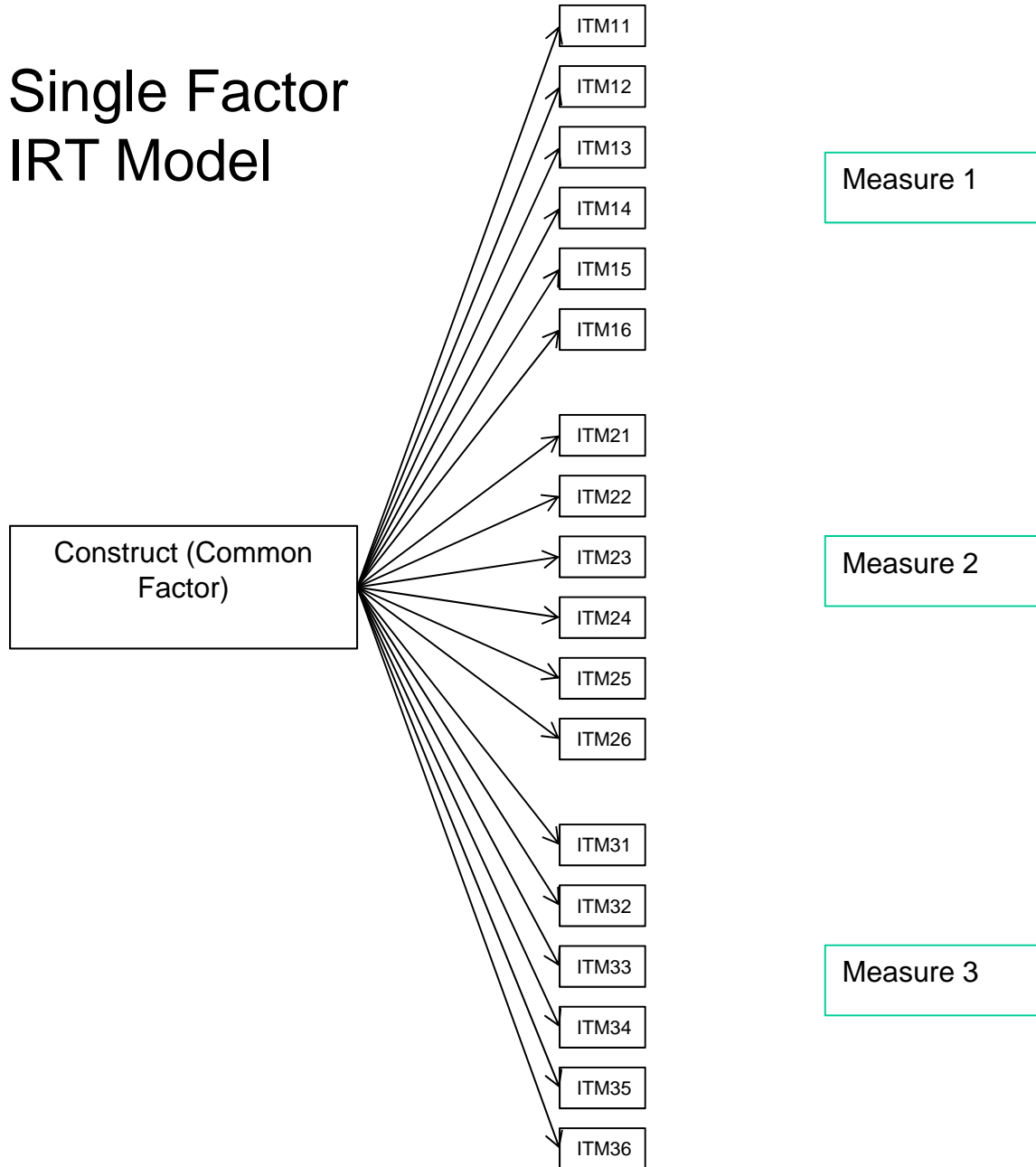
12 Trials  
5 measures

| Trials | Measures |   |   |   |   |
|--------|----------|---|---|---|---|
|        | a        | b | c | d | e |
| 1      | X        | X |   |   |   |
| 2      | X        | X |   |   |   |
| 3      |          |   | X | X | X |
| 4      |          |   | X |   | X |
| 5      |          |   | X |   | X |
| 6      |          |   |   | X | X |
| 7      |          |   |   | X | X |
| 8      |          |   | X |   |   |
| 9      |          |   |   | X |   |
| 10     |          |   |   | X |   |
| 11     |          |   |   | X |   |
| 12     | X        |   | X |   |   |



|            |    |
|------------|----|
| Cells      | 21 |
| Links      | 10 |
| Link Types | 5  |
| Cliques    | 1  |

# Single Factor IRT Model





## Multilevel IRT

- Combine data across all trials (many missing data cells)
- IRT model: single factor loading on all items
- Items are defined as ordered categorical variables
- Multilevel model (clustered within trial)
- Used Full Information Maximum Likelihood estimation, to handle missing data



# IRT Loadings for CDS data

(First five loadings  
For each measure)

| Total N | Trials | ITEM    | PARM         |
|---------|--------|---------|--------------|
| 478     | 3      | CESD001 | <b>0.349</b> |
|         |        | CESD002 | 0.244        |
|         |        | CESD003 | <b>0.277</b> |
|         |        | CESD004 | -0.135       |
|         |        | CESD005 | <b>0.379</b> |
| 98      | 2      | CDRS001 | 0.522        |
|         |        | CDRS002 | <b>1.247</b> |
|         |        | CDRS003 | <b>1.134</b> |
|         |        | CDRS004 | 0.869        |
|         |        | CDRS005 | 0.878        |
| 1291    | 6      | CDI001  | <b>0.43</b>  |
|         |        | CDI002  | <b>1.529</b> |
|         |        | CDI003  | <b>0.56</b>  |
|         |        | CDI004  | <b>0.342</b> |
|         |        | CDI005  | <b>1.938</b> |
| 1850    | 5      | YSR005  | <b>0.804</b> |
|         |        | YSR008  | <b>0.929</b> |
|         |        | YSR012  | <b>1.7</b>   |
|         |        | YSR014  | <b>1.566</b> |
|         |        | YSR018  | <b>2.162</b> |
| 1334    | 4      | CBC008  | <b>0.879</b> |
|         |        | CBC012  | <b>1.073</b> |
|         |        | CBC014  | <b>1.188</b> |
|         |        | CBC018  | <b>1.781</b> |
|         |        | CBC024  | <b>0.638</b> |

**Loadings in red  
significant at  $p < .01$**

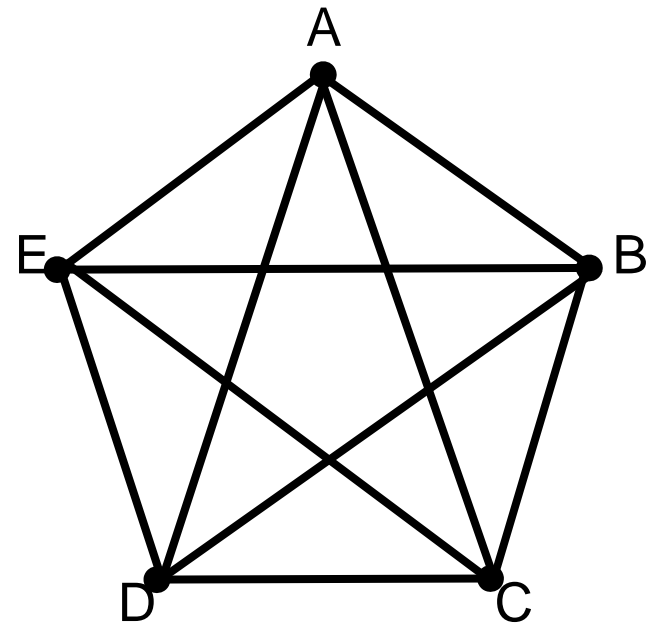


# How will this work with different data structures?

- Continuum of possibilities
  - Saturated matrix: all studies use same measures
  - Partial overlap: studies use different measures with some overlap within some studies
  - No overlap: studies use different measures with no overlap of specific measures across trials

# Saturated Network

|       |   | Measures |   |   |   |   |
|-------|---|----------|---|---|---|---|
| Trial |   | a        | b | c | d | e |
| 1     | X | X        | X | X | X | X |
| 2     | X | X        | X | X | X | X |
| 3     | X | X        | X | X | X | X |
| 4     | X | X        | X | X | X | X |
| 5     | X | X        | X | X | X | X |
| 6     | X | X        | X | X | X | X |
| 7     | X | X        | X | X | X | X |
| 8     | X | X        | X | X | X | X |
| 9     | X | X        | X | X | X | X |
| 10    | X | X        | X | X | X | X |
| 11    | X | X        | X | X | X | X |
| 12    | X | X        | X | X | X | X |

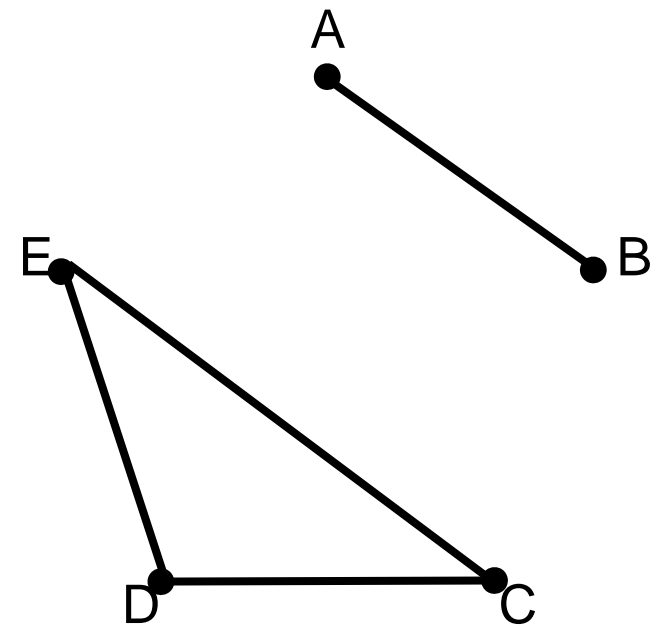


|            |     |
|------------|-----|
| Cells      | 60  |
| Links      | 120 |
| Link Types | 10  |
| Cliques    | 1   |

# Two Clique (2-C):

Densest

| Trials | Measures |   |   |   |   |
|--------|----------|---|---|---|---|
|        | a        | b | c | d | e |
| 1      | X        | X | X |   |   |
| 2      | X        | X | X |   |   |
| 3      | X        | X | X |   |   |
| 4      | X        | X | X |   |   |
| 5      | X        | X | X |   |   |
| 6      | X        | X | X |   |   |
| 7      |          |   |   | X | X |
| 8      |          |   |   | X | X |
| 9      |          |   |   | X | X |
| 10     |          |   |   | X | X |
| 11     |          |   |   | X | X |
| 12     |          |   |   | X | X |

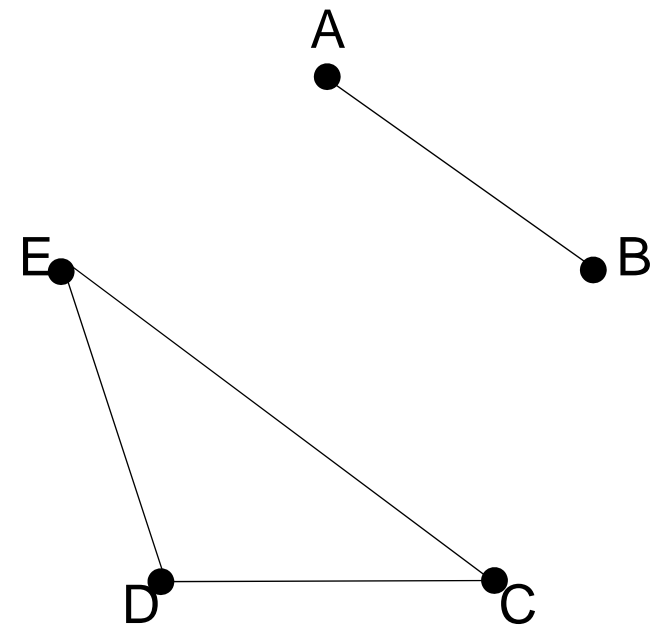


|            |    |
|------------|----|
| Cells      | 30 |
| Links      | 24 |
| Link Types | 4  |
| Cliques    | 2  |

# Two Clique (2-C):

Sparsest

| Trials | Measures |   |   |   |   |
|--------|----------|---|---|---|---|
|        | a        | b | c | d | e |
| 1      | X        | X | X |   |   |
| 2      | X        |   |   |   |   |
| 3      | X        |   |   |   |   |
| 4      |          | X |   |   |   |
| 5      |          | X |   |   |   |
| 6      |          |   | X |   |   |
| 7      |          |   | X |   |   |
| 8      |          |   |   | X | X |
| 9      |          |   |   | X |   |
| 10     |          |   |   | X |   |
| 11     |          |   |   |   | X |
| 12     |          |   |   |   | X |

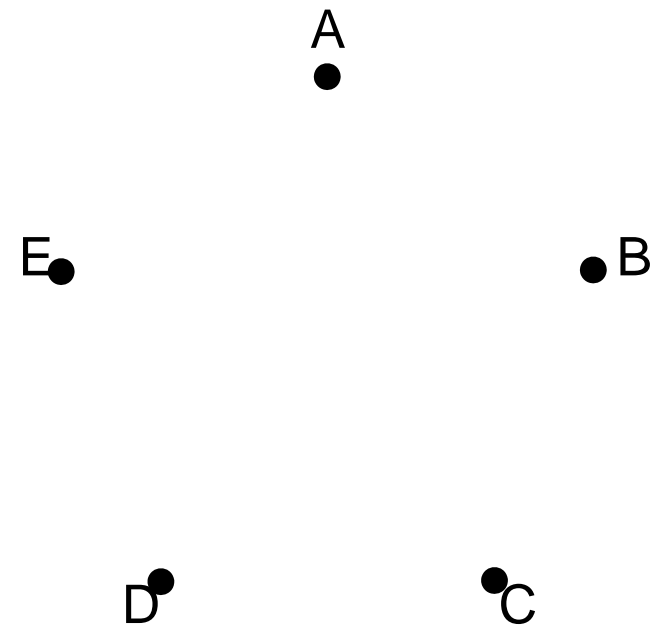


|            |    |
|------------|----|
| Cells      | 15 |
| Links      | 4  |
| Link Types | 4  |
| Cliques    | 2  |

# Five Clique (5-C):

Sparsest

|       | Measures |   |   |   |   |
|-------|----------|---|---|---|---|
| Trial | a        | b | c | d | e |
| 1     | X        |   |   |   |   |
| 2     | X        |   |   |   |   |
| 3     | X        |   |   |   |   |
| 4     |          | X |   |   |   |
| 5     |          | X |   |   |   |
| 6     |          | X |   |   |   |
| 7     |          |   | X |   |   |
| 8     |          |   | X |   |   |
| 9     |          |   |   | X |   |
| 10    |          |   |   | X |   |
| 11    |          |   |   |   | X |
| 12    |          |   |   |   | X |



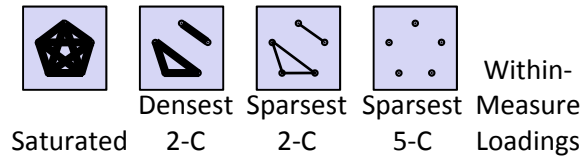
|            |    |
|------------|----|
| Cells      | 12 |
| Links      | 0  |
| Link Types | 0  |
| Cliques    | 5  |

# Simulation study to test IRT model with different data structures

- Used model parameters from analysis of CDS data to construct simulated data (measures strongly inter-correlated)
- Saturated dataset (all trials have all measures)
- “Punched holes” in this dataset to create three other data structures
  - Densest 2-Clique
  - Sparsest 2-Clique
  - Sparsest 5-Clique
- Attempted to estimate FIML IRT for all 4 datasets
- Also estimated within-measure IRTs for each dataset

# IRT Loadings for simulated data

(Based on CDS model: correlated measures)



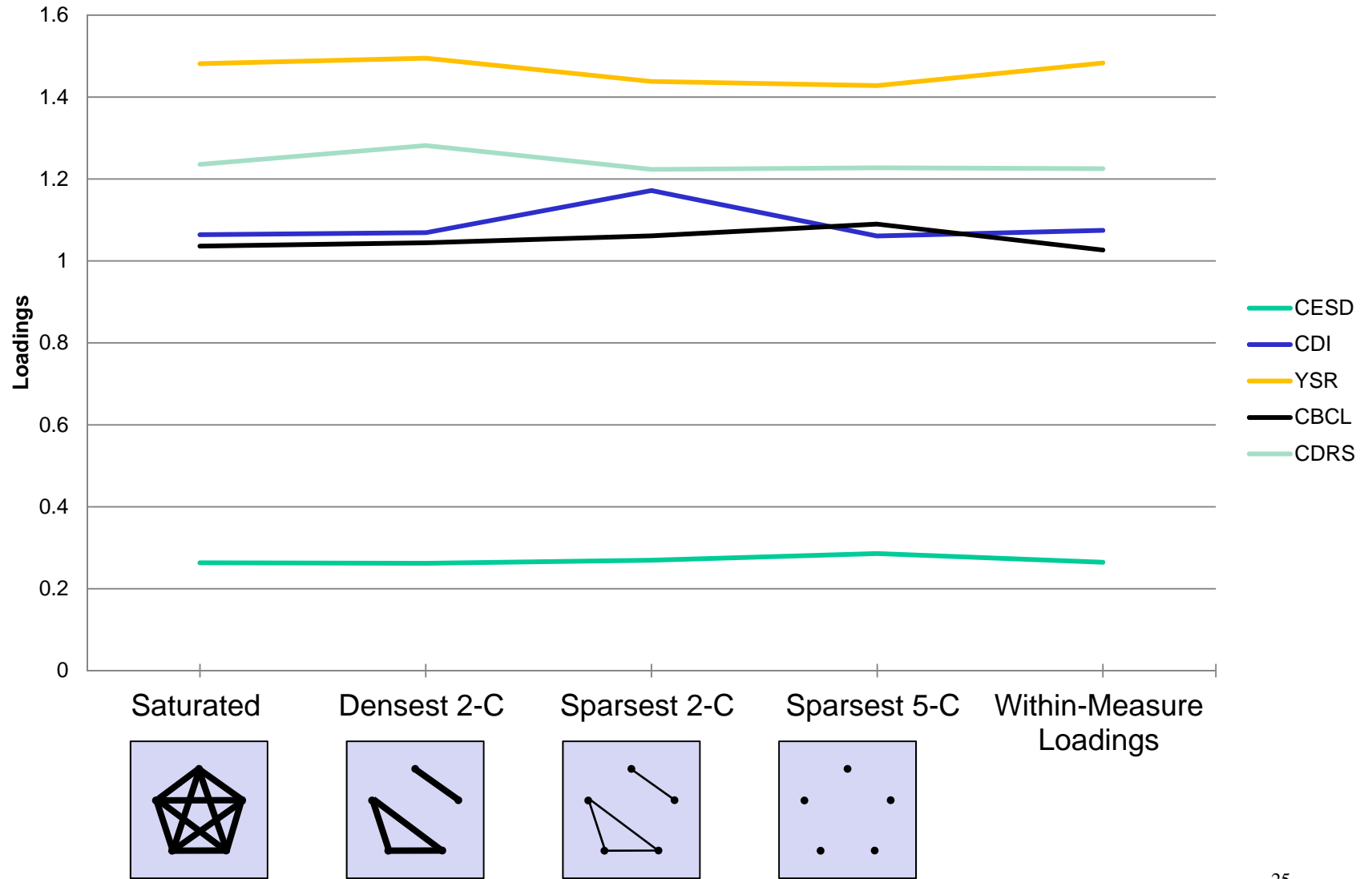
(First five loadings  
For each measure)

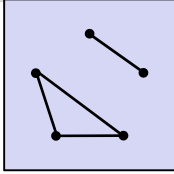
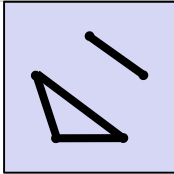
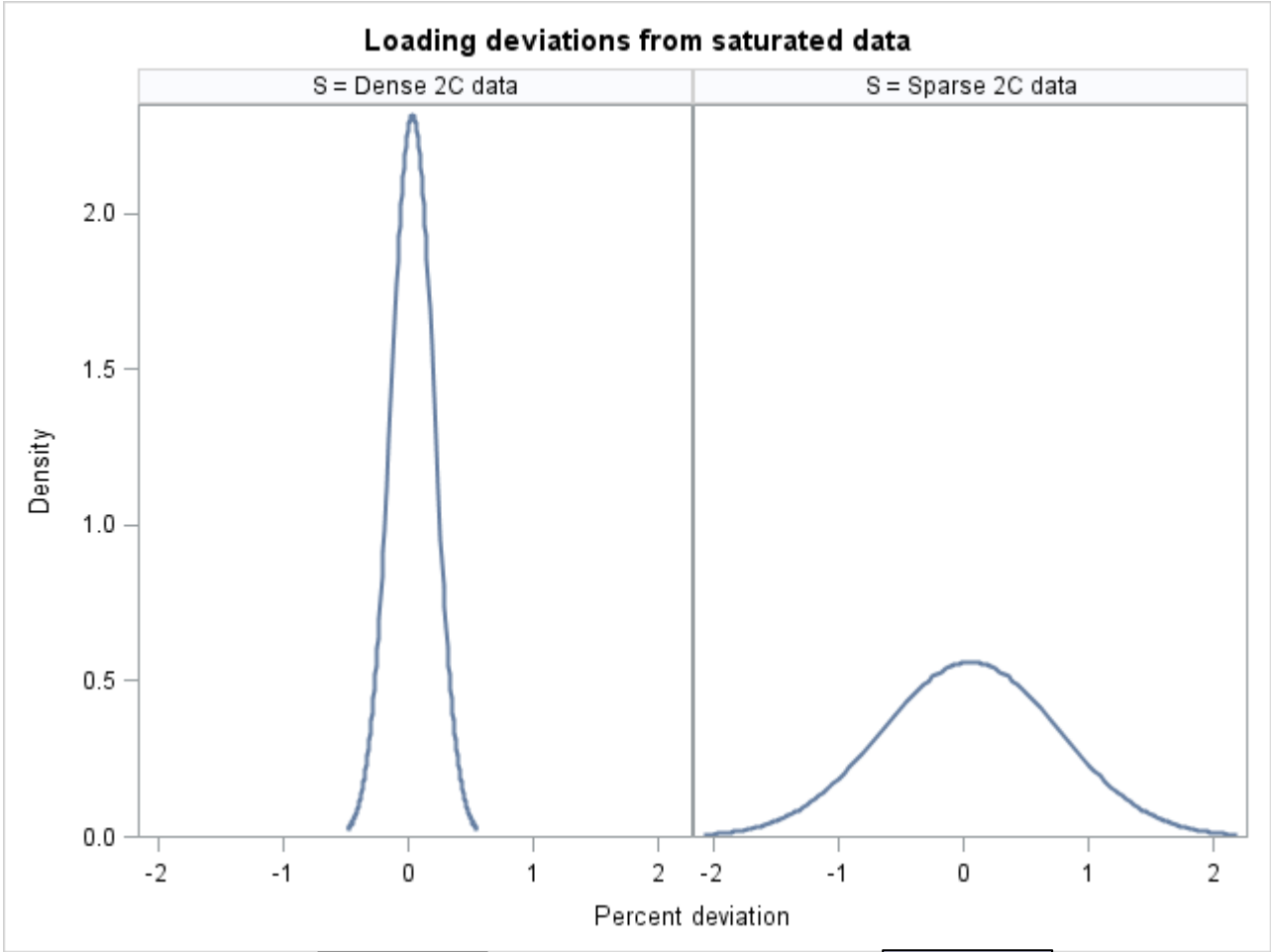
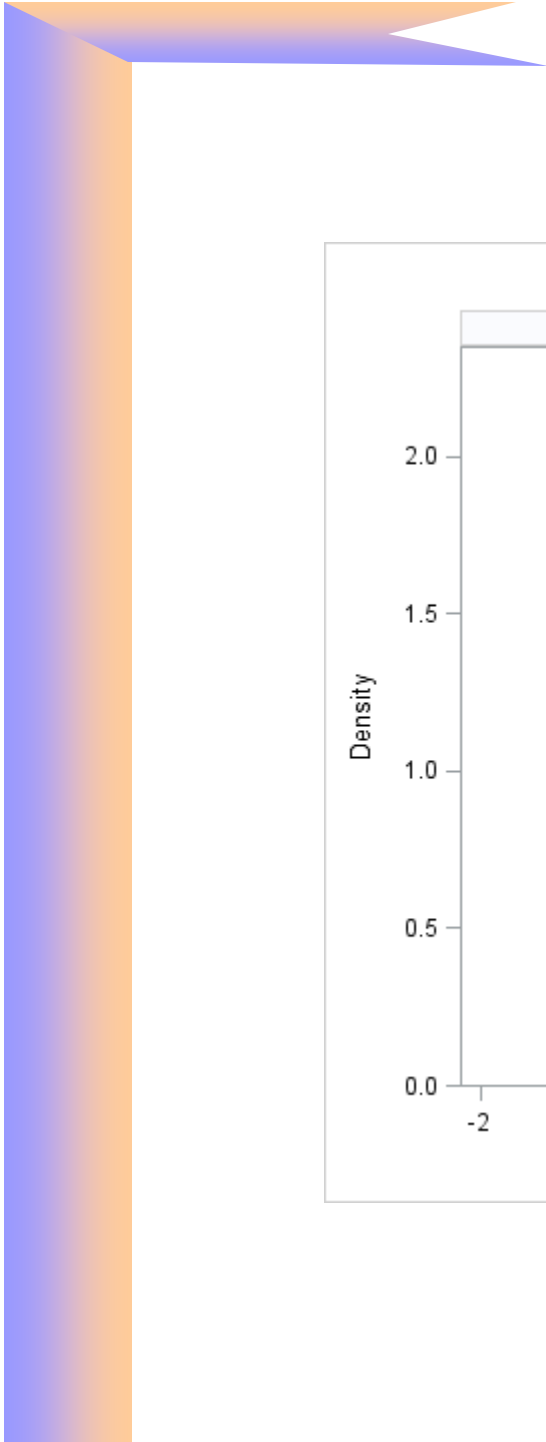
| Measure  | Saturated | Densest 2-C | Sparsest 2-C | Sparsest 5-C | Within-Measure Loadings |
|----------|-----------|-------------|--------------|--------------|-------------------------|
| CESD001  | 0.364     | 0.413       | 0.369        | 0.354        | 0.457                   |
| CESD002  | 0.31      | 0.317       | 0.432        | 0.496        | 0.356                   |
| CESD003  | 0.253     | 0.213       | 0.302        | 0.342        | 0.324                   |
| CESD004  | -0.142    | -0.095      | -0.098       | -0.098       | -0.113                  |
| CESD005  | 0.448     | 0.424       | 0.296        | 0.263        | 0.385                   |
| CDI001   | 0.486     | 0.435       | 0.408        | 0.362        | 0.492                   |
| CDI002   | 1.623     | 1.653       | 1.668        | 1.595        | 1.582                   |
| CDI003   | 0.585     | 0.565       | 0.464        | 0.642        | 0.579                   |
| CDI004   | 0.383     | 0.361       | 0.382        | 0.363        | 0.395                   |
| CDI005   | 2.046     | 1.972       | 1.808        | 1.869        | 2.083                   |
| YSR005   | 0.735     | 0.73        | 0.632        | 0.688        | 0.726                   |
| YSR008   | 0.96      | 0.954       | 0.943        | 0.957        | 0.969                   |
| YSR012   | 1.731     | 1.801       | 1.558        | 1.619        | 1.768                   |
| YSR014   | 1.554     | 1.557       | 1.577        | 1.563        | 1.559                   |
| YSR018   | 2.471     | 2.37        | 2.573        | 2.579        | 2.462                   |
| CBC008   | 0.85      | 0.904       | 0.841        | 0.909        | 0.878                   |
| CBC012   | 1.014     | 1.023       | 1.072        | 1.124        | 1                       |
| CBC014   | 1.152     | 1.137       | 1.251        | 1.276        | 1.171                   |
| CBC018   | 1.627     | 1.725       | 1.144        | 1.27         | 1.53                    |
| CBC024   | 0.648     | 0.662       | 0.706        | 0.665        | 0.662                   |
| CDRSR001 | 0.516     | 0.579       | 0.405        | 0.421        | 0.502                   |
| CDRSR002 | 1.303     | 1.261       | 1.31         | 1.315        | 1.325                   |
| CDRSR003 | 1.223     | 1.297       | 1.159        | 1.157        | 1.211                   |
| CDRSR004 | 0.847     | 0.77        | 0.827        | 0.844        | 0.86                    |
| CDRSR005 | 0.845     | 0.798       | 0.865        | 0.854        | 0.829                   |

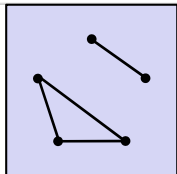
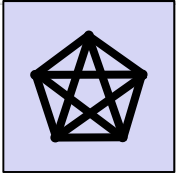
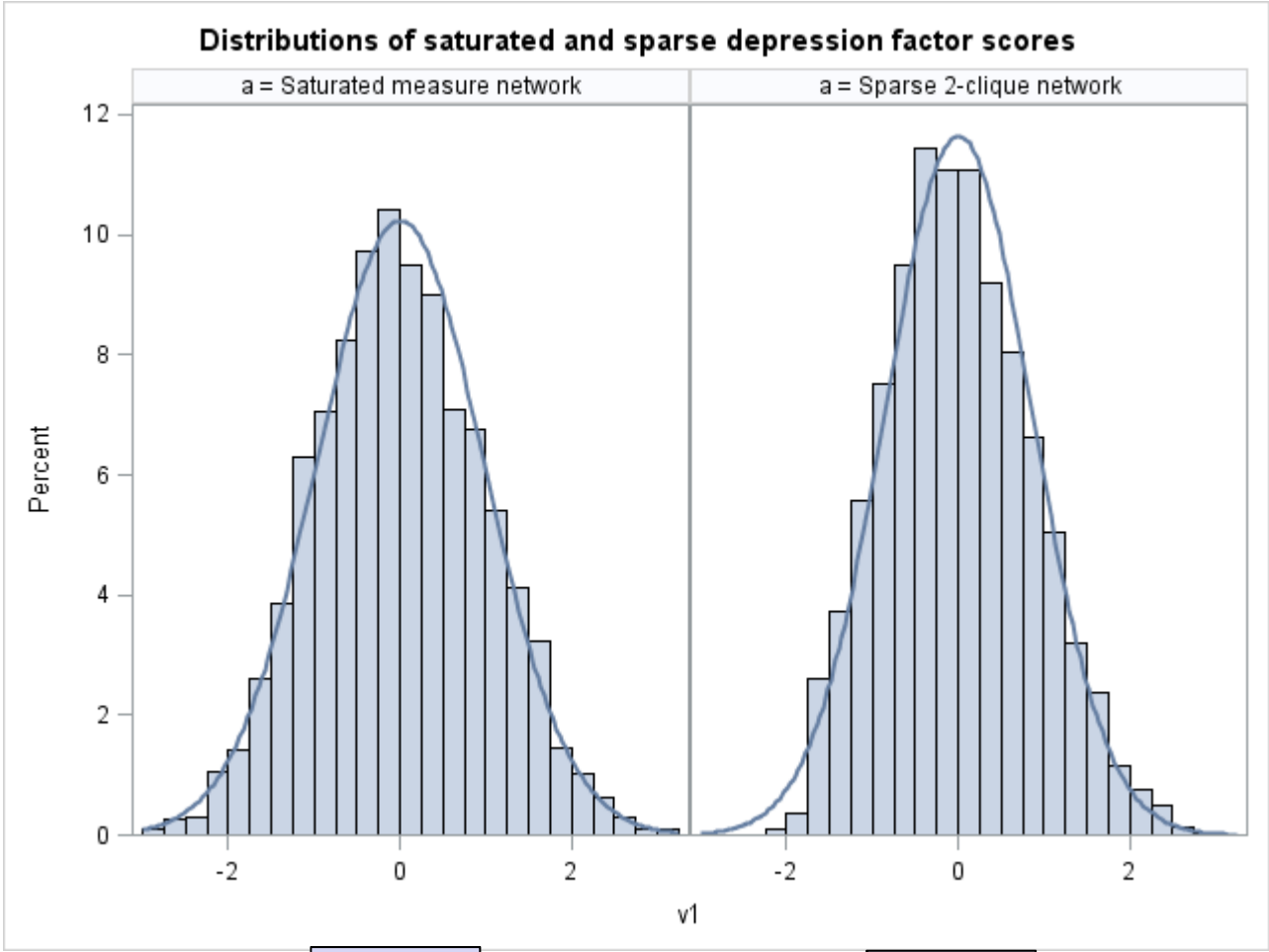
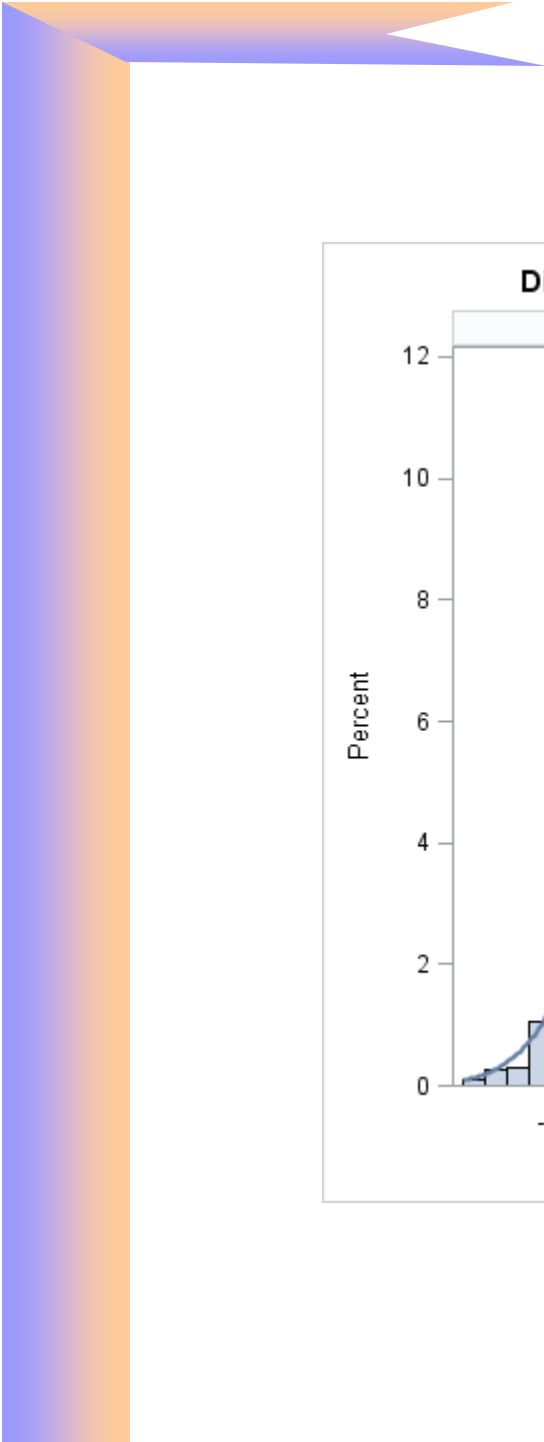
Loadings in red  
significant at  $p < .01$

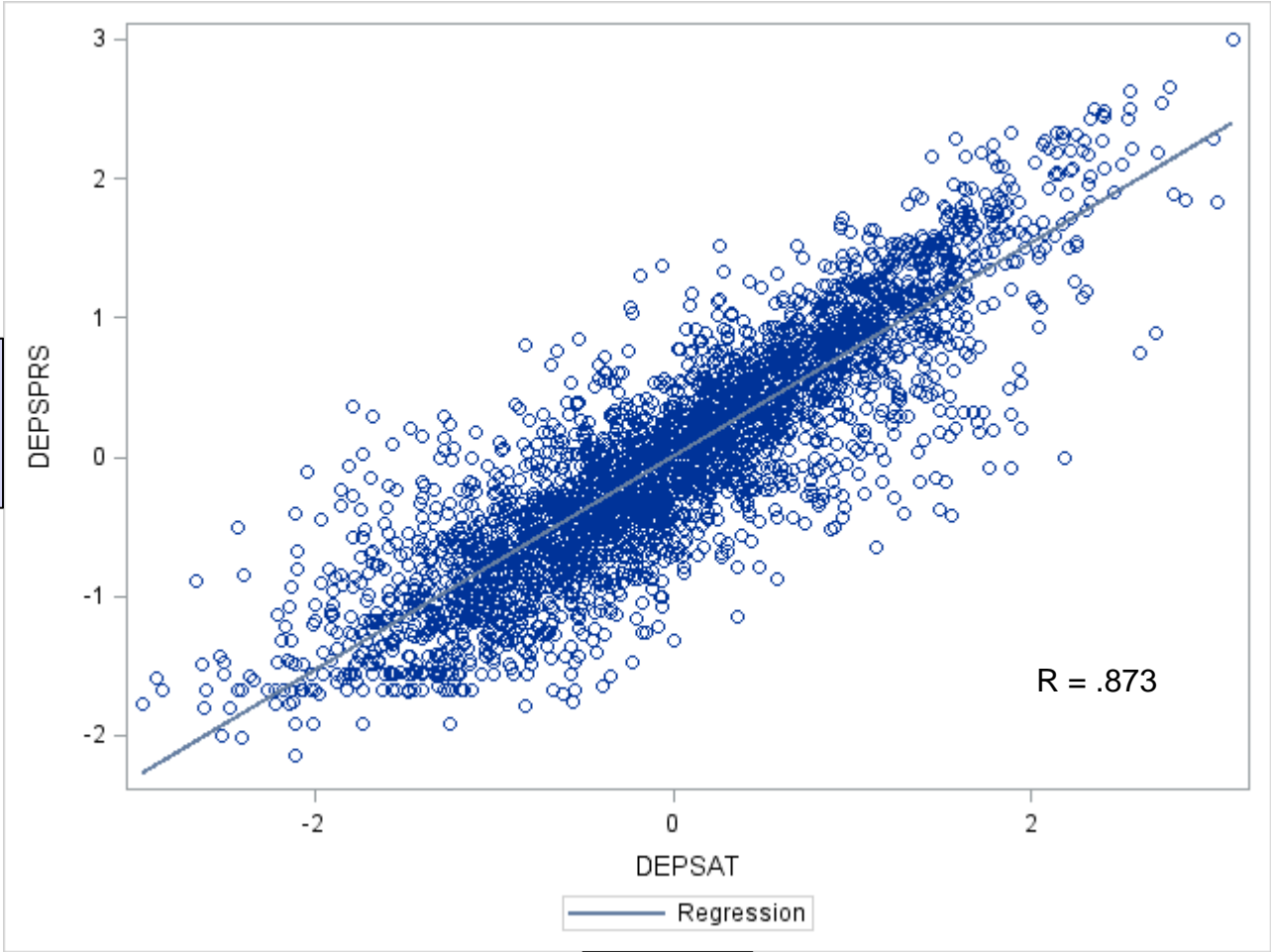
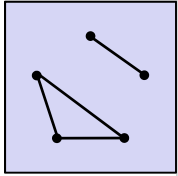
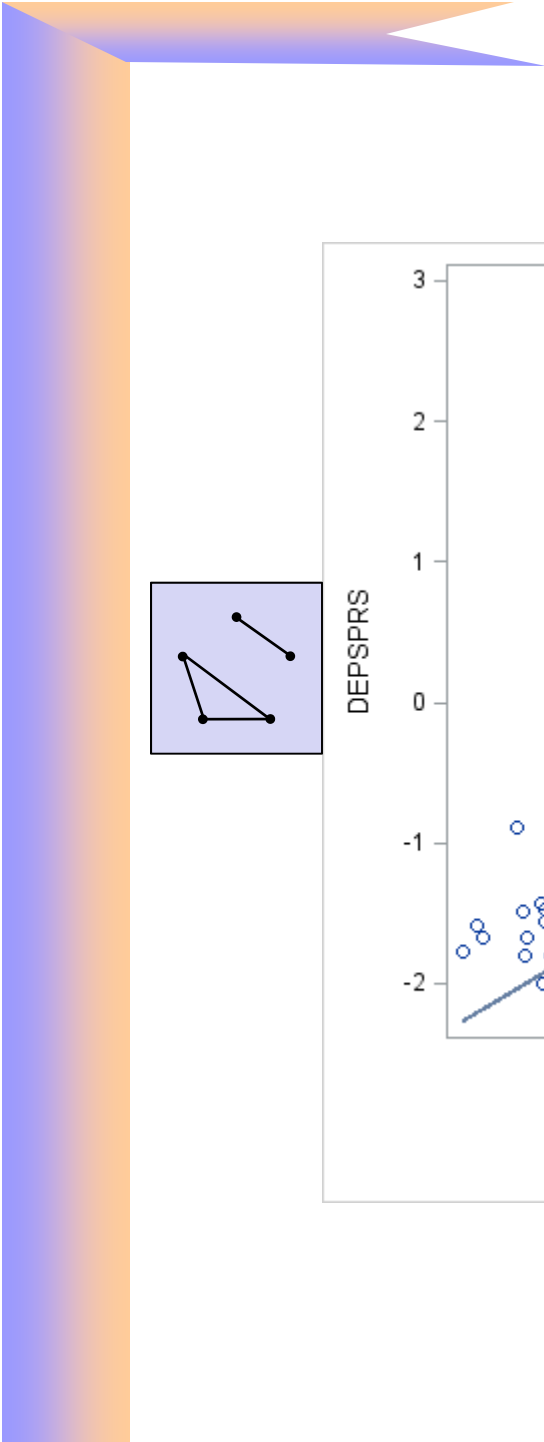


# Mean Item Loadings (simulated data, measures correlated)











## Summary

- Models can be estimated, even when trials have only one measure
- When measures are strongly correlated, IRT returns similar results for loadings regardless of degree of “cliquing”
- Fewer measures increase variability of estimates



## What if measures are not actually correlated?

- Used CDS model to construct new saturated dataset, but forced all measures to be uncorrelated
- Used this to construct 3 datasets with holes
- Repeated IRT models

# IRT Loadings for simulated data (uncorrelated measures)



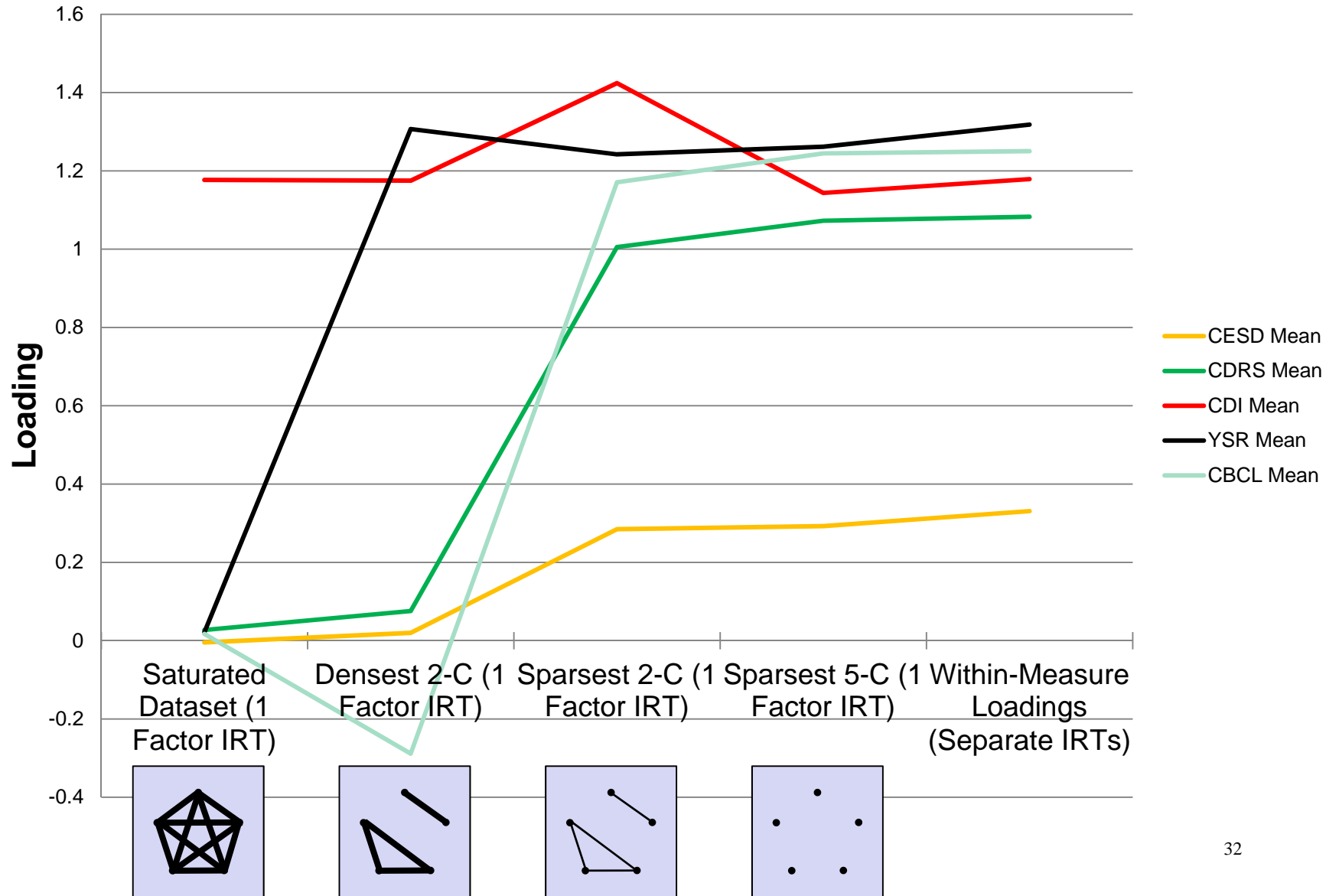
Within-  
Measure  
Loadings

|  | Saturated    | Densest<br>2-C | Sparsest<br>2-C | Sparsest<br>5-C | Within-<br>Measure<br>Loadings |
|--|--------------|----------------|-----------------|-----------------|--------------------------------|
| Loadings   |              |                |                 |                 |                                |
| CESD001  | -0.03        | 0.007          | 0.168           | 0.151           | <b>0.343</b>                   |
| CESD002  | 0.025        | 0.073          | <b>0.356</b>    | <b>0.342</b>    | <b>0.254</b>                   |
| CESD003  | 0.019        | 0.085          | <b>0.398</b>    | <b>0.444</b>    | <b>0.38</b>                    |
| CESD004  | -0.067       | -0.035         | -0.238          | -0.279          | -0.066                         |
| CESD005  | 0.034        | 0.004          | <b>0.439</b>    | <b>0.454</b>    | <b>0.546</b>                   |
| CDRSR001   | 0.035        | 0.099          | <b>0.319</b>    | <b>0.341</b>    | <b>0.344</b>                   |
| CDRSR002   | 0.035        | 0.178          | <b>1.088</b>    | <b>1.105</b>    | <b>1.126</b>                   |
| CDRSR003   | -0.014       | 0.047          | <b>1.157</b>    | <b>1.222</b>    | <b>1.122</b>                   |
| CDRSR004   | 0.002        | 0.001          | <b>0.7</b>      | <b>0.719</b>    | <b>0.738</b>                   |
| CDRSR005   | 0.036        | 0.122          | <b>0.784</b>    | <b>0.813</b>    | <b>0.781</b>                   |
| (First five loadings<br>For each measure) CDI001 | <b>0.505</b> | <b>0.55</b>    | 0.187           | 0.226           | <b>0.509</b>                   |
| CDI002   | <b>1.708</b> | <b>1.561</b>   | <b>1.868</b>    | <b>1.616</b>    | <b>1.707</b>                   |
| CDI003   | <b>0.575</b> | <b>0.57</b>    | <b>0.823</b>    | <b>0.62</b>     | <b>0.573</b>                   |
| CDI004   | <b>0.382</b> | <b>0.279</b>   | 0.202           | -0.015          | <b>0.384</b>                   |
| CDI005   | <b>2.423</b> | <b>2.229</b>   | <b>2.429</b>    | <b>2.362</b>    | <b>2.426</b>                   |
| YSR1   | 0.013        | <b>1.746</b>   | <b>1.599</b>    | <b>1.529</b>    | <b>1.776</b>                   |
| YSR2   | -0.025       | <b>0.17</b>    | <b>0.417</b>    | <b>0.51</b>     | <b>0.212</b>                   |
| YSR3   | -0.043       | <b>0.455</b>   | 0.293           | <b>0.38</b>     | <b>0.455</b>                   |
| YSR4   | 0.015        | <b>1.58</b>    | <b>1.778</b>    | <b>1.663</b>    | <b>1.667</b>                   |
| YSR5   | 0.057        | <b>2.712</b>   | <b>2.38</b>     | <b>2.425</b>    | <b>2.786</b>                   |
| CBC1   | -0.052       | -0.172         | <b>1.123</b>    | <b>1.154</b>    | <b>0.899</b>                   |
| CBC2   | -0.024       | -0.144         | <b>0.514</b>    | <b>0.536</b>    | <b>0.376</b>                   |
| CBC3   | 0.014        | -0.448         | <b>1.459</b>    | <b>1.585</b>    | <b>1.679</b>                   |
| CBC4   | 0.015        | -0.226         | <b>0.998</b>    | <b>0.985</b>    | <b>1.184</b>                   |
| CBC5   | 0.033        | -0.275         | <b>1.074</b>    | <b>1.072</b>    | <b>1.15</b>                    |

Loadings in red  
significant at  $p < .01$

# Mean Item Loadings

(Simulated data, measures uncorrelated)

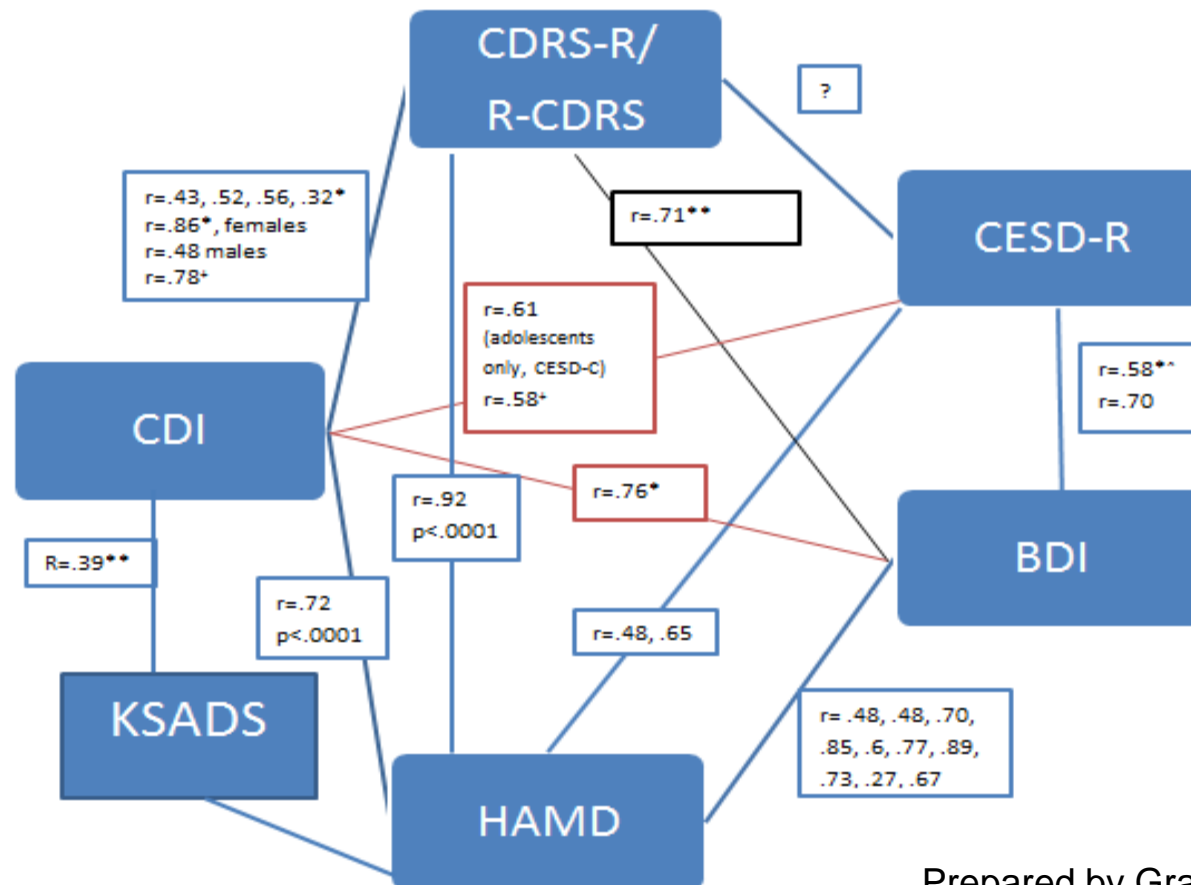




## Summary

- Factor scores generated from IRT may be very useful way of establishing “harmonized” measure, but diagnostics are essential to determine whether model is reasonable
- When measures are uncorrelated (no single construct), IRT appears to be driven by internal consistency of measures
  - Saturated data: selects measure with best item inter-correlations
  - Seems to do the same thing for each clique
  - The 5-clique model is similar to meta-analysis: each measure is a separate index
- Even with complete isolation (5 cliques), estimates are stronger because they are based on all trials with that measure.
- Adding information about measure correlations from other studies may be useful way of strengthening estimates from highly isolated datasets, even if we don’t have individual-level data

# Correlations across depressive symptom measures from other literature



Prepared by Gracelyn Cruden

+significant differences found by gender and/or race, age  
 ^adolescent mothers comprise entire sample  
 \* $p < .05$  or  $.01$   
 \*\* $p < .001$



## Future Directions

- Harmonizing existing data
  - Including measure correlations from other studies
  - Adding bridge measures to existing studies
  - Testing DIF across trials (may require “parceling” items)
- Developing measurement systems for next generation trials
  - Multilevel IRT findings may point toward sets of indicators more likely to support construct equivalence
  - Can also support adaptive measurement, where only subset of items are used for any respondent