Small-area estimation of prevalence of serious emotional disturbance (SED) in schools

Alan Zaslavsky
Harvard Medical School
Overview

• Detailed domain data from short scale
• Limited amount of data from calibration survey with longer instrument
• Models relating the instruments at individual and school levels
• Predictions based on data from short scale
  – School level (prevalence)
  – Individual level (screening)
Outline

• Measures
  – CIDI
  – K6 and enhancement for adolescents

• Sample design

• Models

• Predictions

• Future directions
Measures: CIDI

• Composite International Diagnostic Interview
  – Trained lay interviewers
• Adolescent version: CIDI-A
  – Adolescent interview
  – Parent questionnaire (SAQ) focused on dx more noticed by parents (ADHD, )
• Contains most information required for SED dx
  – P(Clinical GAS) imputed from CIDI dx & other items (data from 347 clinical validation interviews)
• Numerous other covariates
Measures: K6 scale

• 6 items on 0-4 severity scale (never–always)
  – “So depressed nothing could cheer you up”
  – Nervous
  – Restless/fidgety
  – Hopeless
  – Everything an effort
  – Worthless

• Internalizing disorders: depression, anxiety
  – Developed for adult population
K6: adolescent enhancements

- Considered 18 items: screeners, behavioral symptoms of personality disorder
- Elicit externalizing disorders (ODD, CD, ADHD)
  - Earlier onset, more common in adolescents
- Selected 5 items
  - Screeners for ADHD, IED, ODD
  - 2 personality items: “can stay out of trouble”, “have strong temper”
K6: predictive power at individual level

- Compare by AUC (area under ROC curve) for prediction of disorders and SED
  - 0.5=chance agreement, 1.0=perfect agreement

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>K6</th>
<th>K6+5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any mood disorder</td>
<td>.77</td>
<td>.81</td>
</tr>
<tr>
<td>Any anxiety disorder</td>
<td>.73</td>
<td>.75</td>
</tr>
<tr>
<td>Any behavior disorder</td>
<td>.67</td>
<td>.82</td>
</tr>
<tr>
<td>Any SED</td>
<td>.74</td>
<td>.83</td>
</tr>
<tr>
<td>SED with behavioral disorder</td>
<td>.53</td>
<td>.78</td>
</tr>
</tbody>
</table>
Sample design of NCS-A

- **National Comorbidity Survey – Adolescent Supplement**

- **School-based component**
  - Highly stratified national PPS sample
  - 320 schools (after replacement of refusing schools)

- **Final sample** 282 schools (≥10 students/school)

- **9244 adolescents** (74.7% participation rate)
  - 83.7% parent response (conditional on child)
Models

• Bivariate multilevel mixed model, continuous outcomes $Y_1, Y_2$:

$$Y_{ijm} = X_{ijm} \beta_m + \nu_{im} + \epsilon_{ijm}$$

$i =$ cluster (school, neighborhood, etc.)

$j =$ individual

$m = 1, 2 =$ measure

$X =$ covariates

$\nu =$ cluster-level random effect

$\epsilon =$ individual-level random effect
Models

• Version with dichotomized SED outcome:

\[ Y_{ij3} = I\{Y_{ij2} > 0\} \]

– Equivalent to probit model for binary outcome
– E.g. \(Y_1=\) screener score, \(Y_2=\Phi(P(\text{SED}))\), \(Y_3=\text{SED}\)

• Covariates

– Age, sex, race/ethnicity, age at school entrance
– Public/private, size (<50 teachers vs >50)
Model diagram

\[ \begin{align*}
\mathbf{X}_{ij1} & \rightarrow \mathbf{Y}_{ij1} & \beta_1 \\
\mathbf{X}_{ij2} & \rightarrow \mathbf{Y}_{ij2} & \beta_2 \\
\epsilon_{ij1} & \leftrightarrow \epsilon_{ij2} & \sum_e \\
U_i1 & \leftrightarrow \sum_v U_i2 
\end{align*} \]
Estimates from NCS-A

- With continuous outcomes

<table>
<thead>
<tr>
<th></th>
<th>$v$ (school level)</th>
<th>$e$ (individual level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_1$ (K6+5 variance)</td>
<td>0.019</td>
<td>0.371</td>
</tr>
<tr>
<td>$\sigma^2_2$ (SED variance)</td>
<td>0.037</td>
<td>0.597</td>
</tr>
<tr>
<td>$P$ (correlation)</td>
<td>0.845</td>
<td>0.544</td>
</tr>
</tbody>
</table>
Covariates

• Age, sex, race-ethnicity
• Age starting school
• Public/private
Out-of-sample prediction

Scenario:
• Collect K6+5 measures in school subsample.
• Predict for remainder of school
Individual-level prediction (naïve)

Scenario:
• Collect K6+5 measures for individuals.
• Predict SED score for same individuals
• Ignore clustering
In-sample prediction

Scenario:
• Collect K6+5 measures for a sample.
• Predict SED score for same sample.
• Design with sampling within school combines in- and out-of-sample prediction
Prediction for individuals

• Special case of in-sample prediction
• Estimate of school-level random effect has substantial effect on prediction for individual.
  – Implies prevalence.
  – E.g. same screening score at schools at ±1 SD from median random effect → P(SED)=12.7%, 6.3% respectively
Limitations and open questions

• Short scale development
  – Only items from current CIDI questionnaire
  – Might function differently out of context
  – Investigate other short scales

• Validation sample design
  – Optimize for estimation of variance/covariance
  – Test school-level covariates

• Model fit imperfect at high end
  – Perhaps better for ranking than exact prevalence
More covariates

• School level: Principal questionnaire
  – Frequency of depression reports, attacks/fights, etc
• Neighborhood characteristics
  – Families in poverty, homeless, racial/ethnic composition, stability, region, urbanicity
• School-level (ecological) models – weaker than K6
  – $R^2=0.45$ with both sets of variables
  – $R^2=0.40$ with only Census
• Need data on school characteristics related to assignment policies/patterns
Conclusion

• Combination of short screening scale with calibration survey is technically possible
• Large improvements relative to synthetic model based on demographics
• Best estimates use multilevel model instead of single-level regression.
References

• NCS-A description

• K6+ development

• Modeling