Latent class models for individual participant data meta-analyses of diagnostic test accuracy studies with imperfect reference standards

Zelalem Negeri, Ph.D.
Postdoctoral Fellow
Department of Epidemiology Biostatistics and Occupational Health
McGill University

Joint work with:
Brooke Levis, Yin Wu, Nandini Dendukuri, Brett Thombs and Andrea Benedetti

A day with ... SMG
May 11, 2021
1. Background – Overview

• The Patient Health Questionnaire-9 (PHQ-9) is a nine-item questionnaire widely used to screening for major depression

• The PHQ-9 has been evaluated against various reference standards:
  - Semi-structured interviews (e.g., SCID)
  - Fully structured interviews (e.g., CIDI, MINI)

• The PHQ-9 database:
  - Accrued by the DEPRESsion Screening Data Project (www.depressd.ca)
  - More than 100 studies comprising of about 44,000 participants and over 4,500 major depression cases
1. Background – The Problem

• The PHQ-9 database:
  ▪ Multiple reference standards (e.g., SCID, CIDI MINI) with different depression diagnosing capabilities
  ▪ A single reference standard per participant

• Wu et al. (2020):
  ▪ Synthesized three individual participant data meta-analyses (IPDMAs) – that included 69,405 participants from 212 studies
  ▪ The MINI categorized major depression more frequently relative to the SCID

• This could influence estimates of PHQ-9 sensitivity, specificity, and depression prevalence
2. Objectives

- We aimed to propose and validate LCMs for the IPDMA of the PHQ-9 to accurately estimate:
  - PHQ-9 sensitivity and specificity
  - Reference standards’ sensitivity and specificity
  - Depression prevalence
3. Methods – Frequentist LCMs

• We propose and validate both Frequentist- and Bayesian-based LCMs by assuming the true depression status as unknown

• Our Frequentist LCM assumes a common sensitivity and specificity across studies but study-specific prevalence

  ▪ We additionally assume conditional dependence between PHQ-9 and reference standards (i.e., FCDLCM)

  ▪ We use the expectation-maximization (EM) algorithm to estimate model parameters
3. Methods – Bayesian LCMs

• We also propose a Bayesian conditional dependence LCM (BCDLCM) by introducing covariance parameters as in the FCDLCM

• The BCDLCM assumes the multinomial distribution to model the observed cell frequencies (TPs, FPs, FNs and TNs) given the unobserved disease status and model parameters

  ▪ We use the Gaussian distribution as priors for the logit-transformed pooled sens, spec, depression prevalence, and random effects; and the Uniform distribution as hyper-priors for the precision parameters

  ▪ We use Markov Chain Monte Carlo (MCMC) as implemented in the R package rjags to sample from the marginal posterior distribution of each model parameter
4. Results – Simulations

Table 1: Results when true PHQ-9 sens=spec = 0.7, $\tau_1^2 = 1.2$, $\tau_2^2 = 0.6$, $\rho = -0.6$, $k = 30$, $n = 134$, $\pi = 10\%$ when data were generated assuming conditional dependence and MINI as imperfect

<table>
<thead>
<tr>
<th>Model</th>
<th>Bias</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PHQ Sens</td>
<td>PHQ Spec</td>
</tr>
<tr>
<td>BREM</td>
<td>-0.22</td>
<td>-0.02</td>
</tr>
<tr>
<td>FCDLCM</td>
<td>-0.07</td>
<td>0.01</td>
</tr>
<tr>
<td>BCDLCM</td>
<td>0.02</td>
<td>0.10</td>
</tr>
</tbody>
</table>

BREM=Bivariate Random-Effects Model; FCDLCM=Frequentist Conditional Dependence Latent Class Model; BCDLCM=Bayesian Conditional Dependence Latent Class Model
4. Results – Real Data

Table 2: Estimates of PHQ-9 and MINI sensitivity and specificity in % when PHQ-9 was compared against the MINI interview at the PHQ-9 standard cut-off of ≥ 10

<table>
<thead>
<tr>
<th>Model</th>
<th>PHQ Sens (95% CI)</th>
<th>PHQ Spec (95% CI)</th>
<th>MINI Sens (95% CI)</th>
<th>MINI Spec (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BREM</td>
<td>74 (67, 79)</td>
<td>89 (86, 91)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>FCDLCM</td>
<td>83 (73, 90)</td>
<td>94 (87, 96)</td>
<td>91 (87, 96)</td>
<td>99 (99, 100)</td>
</tr>
<tr>
<td>BCDLCM</td>
<td>82 (73, 92)</td>
<td>92 (87, 95)</td>
<td>68 (58, 82)</td>
<td>97 (94, 99)</td>
</tr>
</tbody>
</table>

BREM=Bivariate Random-Effects Model; FCDLCM=Frequentist Conditional Dependence Latent Class Model; BCDLCM=Bayesian Conditional Dependence Latent Class Model
5. Conclusions

- We proposed robust models that can handle multiple imperfect reference standards.

- As expected, the BREM performed well when diagnostic interviews were assumed perfect, otherwise underestimated PHQ-9 sensitivity.

- LCMs were shown to be alternative approaches for IPDMA to account for differences in reference standard accuracy.