Assessing Concordance among Diagnostic Tests for Malaria using Latent Class Analysis

Noory Kim
BIOS 841
April 2013

Issue: There is no gold standard diagnostic test for malaria.

Method	Sensitivity	Specificity	Time	Cost
Microscopy	Low when parasite densities are low	Low to Medium	Short to Long: need to have microscope on-site or ship to lab	Expensive: laborious
PCR	Low to High: depends on parasite densities	Low to High	Long: need to ship sample to lab	Expensive: laborious
Rapid diagnostic tests	?	?	Short: on-site	Less expensive

Methodologies for assessing concordance

	Agreement Indices (e.g. kappa statistic)	Latent class analysis
Assesses degree of rater concordance/ discordance	✔	✔
Differentiates between sensitivity and specificity	No	✔
Consistent over changes in prevalence rates	No	?

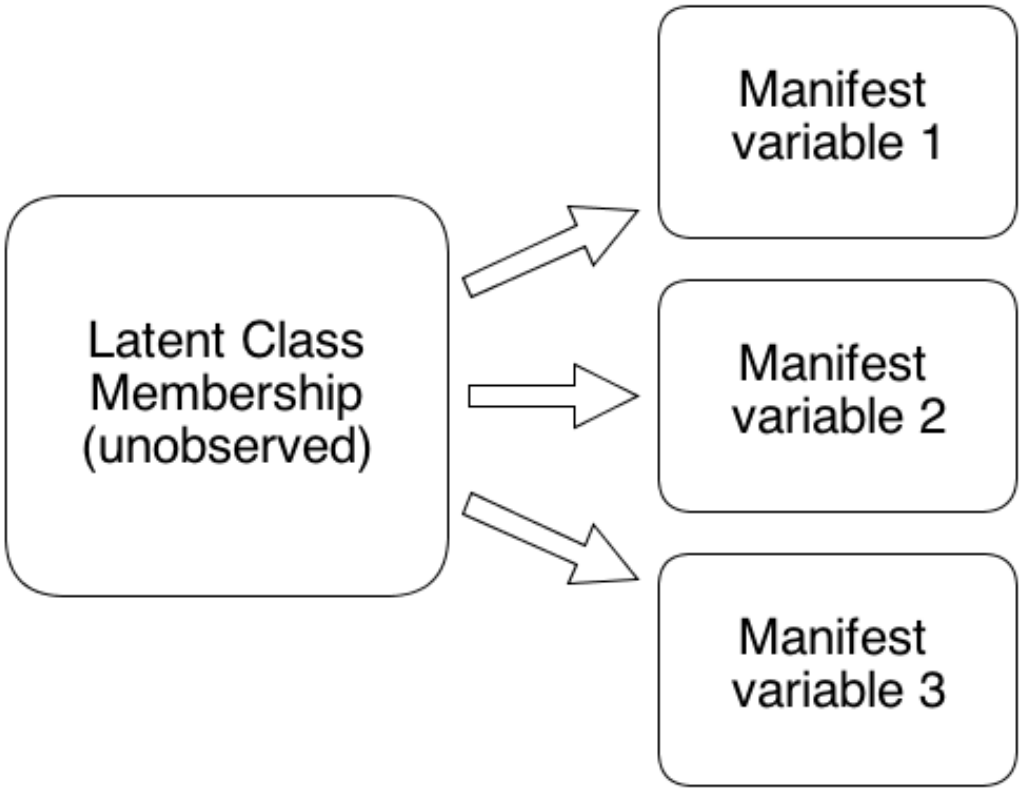
Latent class analysis

- Assumes each individual in a study population belongs to one of several "latent classes"
  - Memberships are not directly observable. Ex. malarial disease status.
  - Memberships are inferred from a combination of observed categorical "manifest variables". Ex. survey responses; different malarial diagnostic tests.
- Typical Goals
  - Quantitative: Determine the lowest number of latent classes that are distinct from each other.
  - Qualitative: See and describe how the classes differ from each other. ("latent class separation")
- Assumptions
  - Conditional independence: Within each latent class, the manifest variables are assumed to be independent.
    - Realistic?

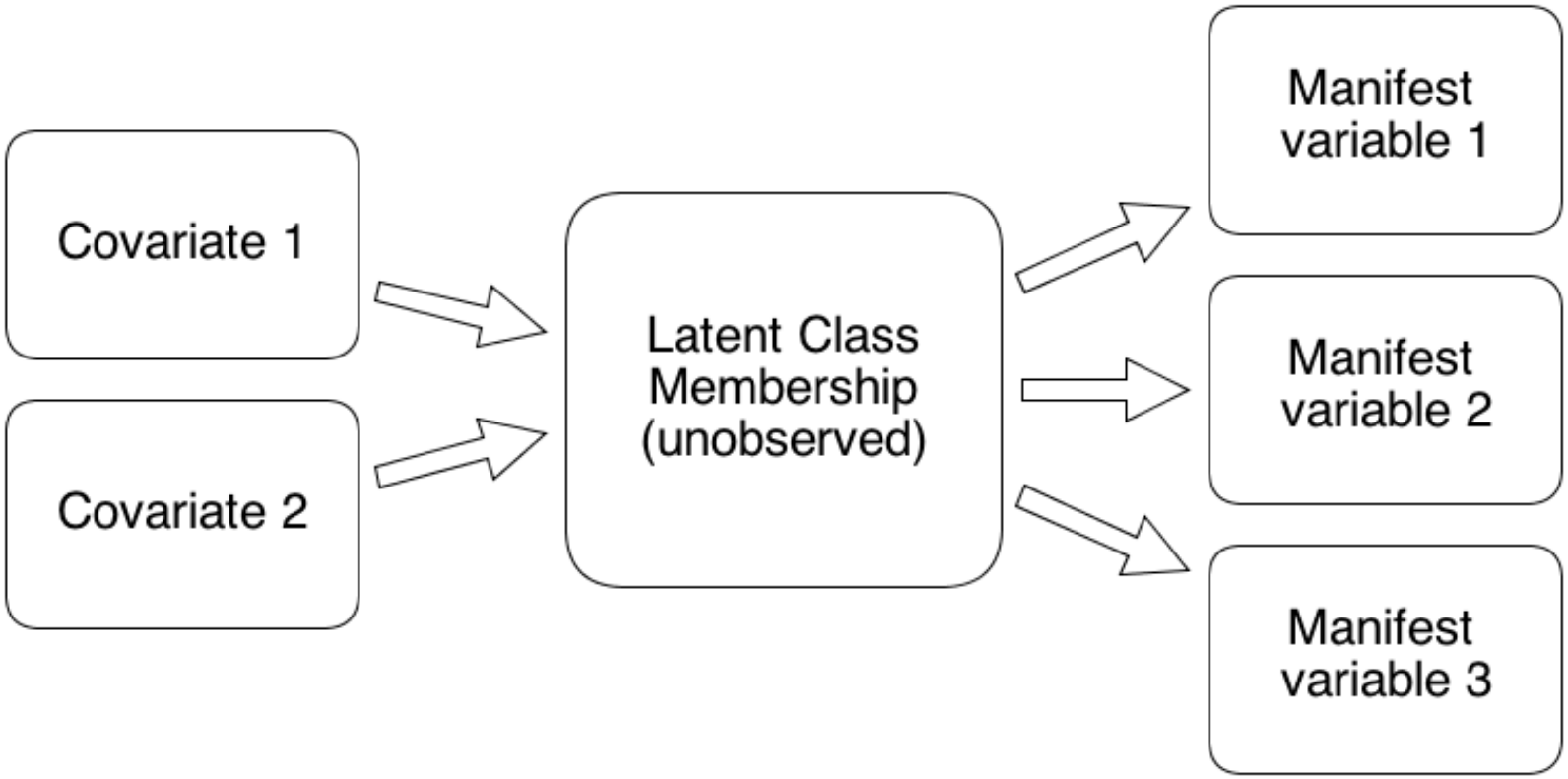
Models Considered

- Goal: To assess degree of concordance between three types of diagnostic tests for malaria
- Models considered
  - Originally in the Statistical Analysis Plan (SAP)
    - Basic latent class model (with no covariates)
    - Latent class model with covariates
      - Age
      - Gravidity
  - Not in the SAP
    - Latent class model with covariates treated as manifest variables
- Software used: R package poLCA (Linzer and Lewis, 2011)

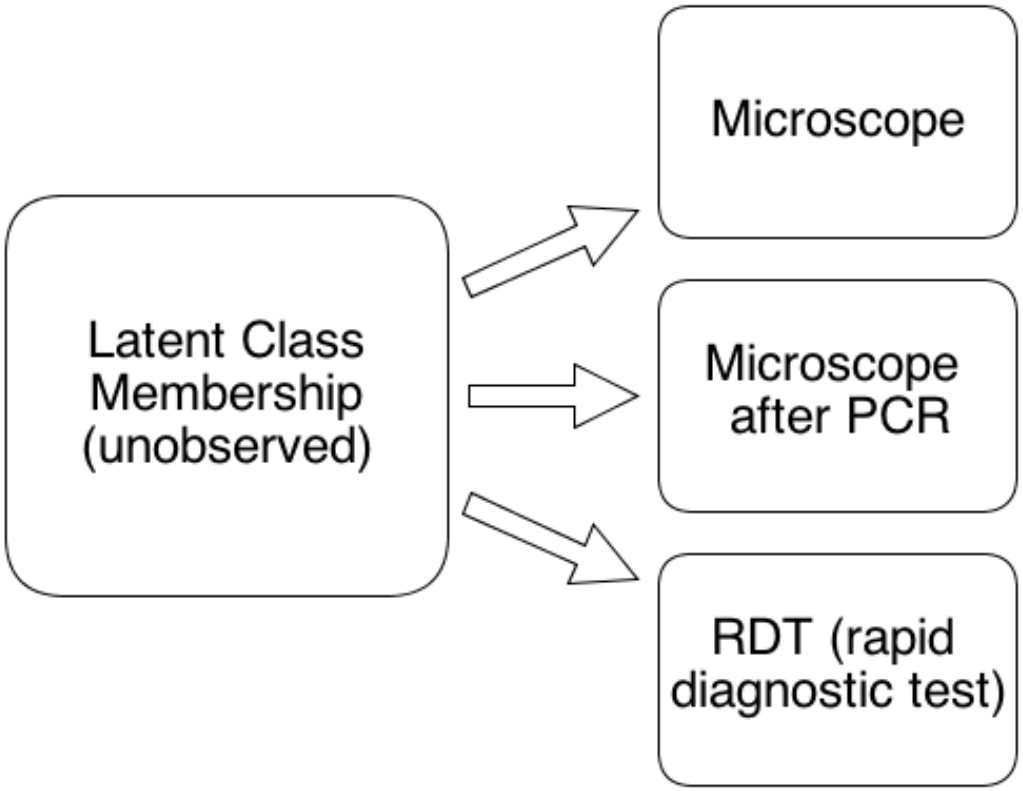
Basic LCA Model



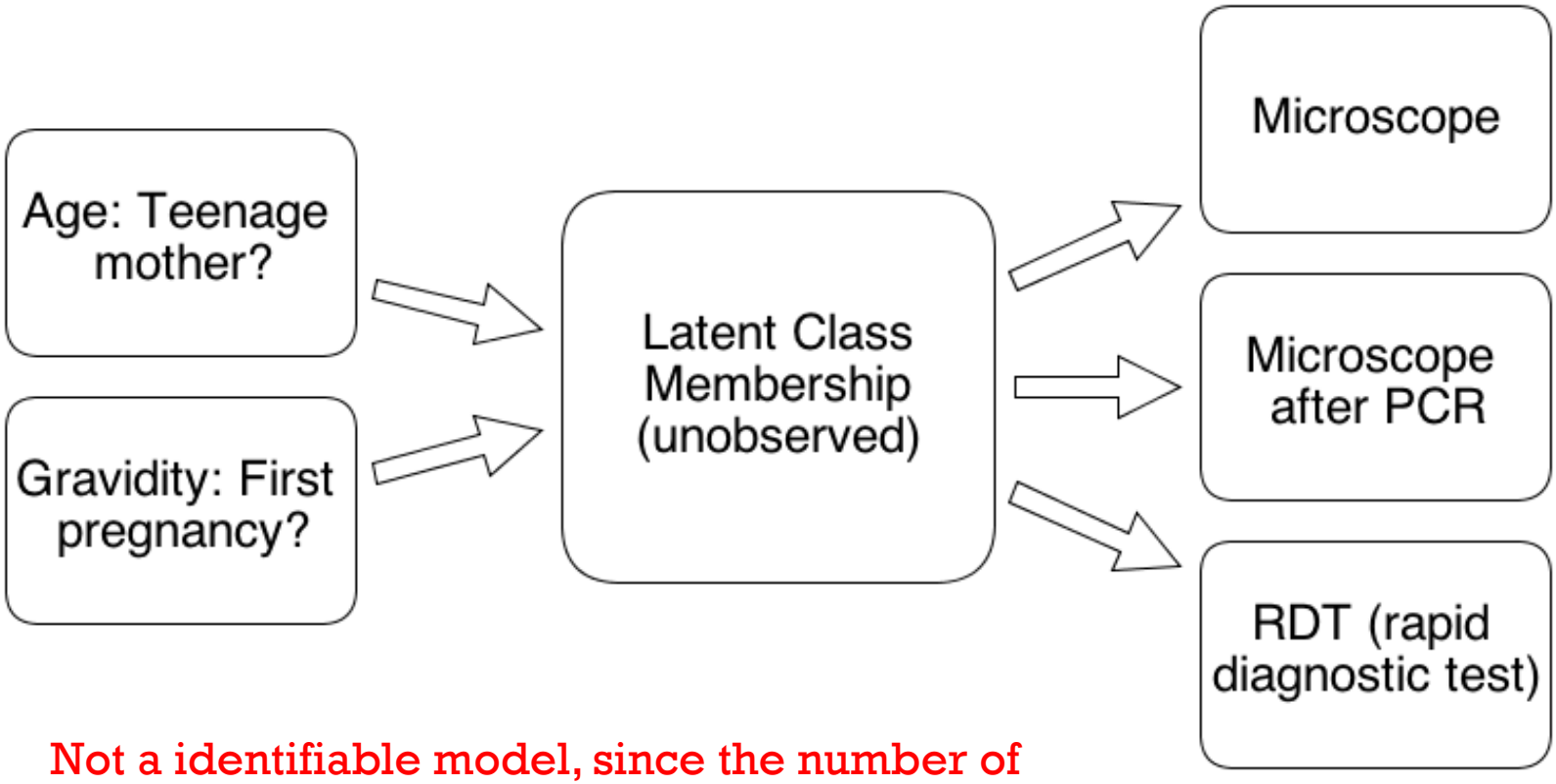
LCA Model with Covariates



Basic LCA Model

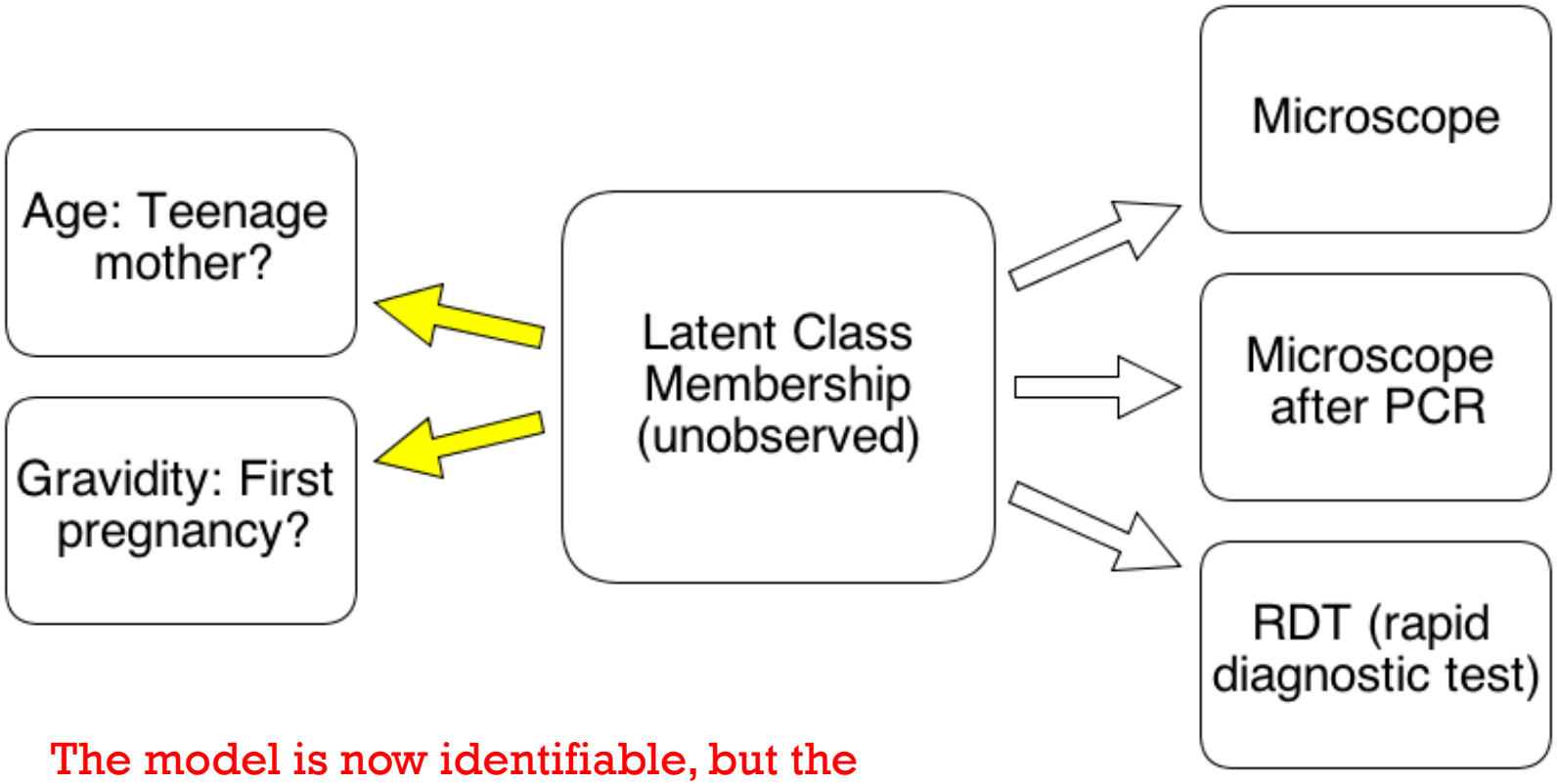


LCA Model with Covariates



Not a identifiable model, since the number of parameters to be estimated exceeds the degrees of freedom (which equals the number of cells in the contingency table minus 1).

Treating Covariates as Manifest Variables



The model is now identifiable, but the interpretability of the covariates has changed.

Data Cleaning & Preparation

- > 5740 records in 4 data tables; merged tables by participant ID number
- Data cleaning
  - Dichotomize diagnostic test results, time, and covariates.
    - Test: 1 if any malaria detected; 0 otherwise (except missing values)
    - Time of test: (*NumberVisit*='mEli') → (*trimester*=2); o.w. (*trimester*=3)
    - Age: (*ScrMotherAge*<20) → (*teen*=1); o.w. (*teen*=0)
    - Gravidity: (*ScrMotherChild*=0) → (*firstpreg*=1); o.w. (*firstpreg*=0)
  - Remove records that are duplicates (by id, trimester, and test result).
  - Set as missing any conflicting results for the same test in records having the same id and same trimester.
- Combine results for same test from different trimesters using a "once positive, always positive" assumption
  - Since no complete cure malaria exists, and
  - Since not all samples from a person with malaria necessarily have malarial parasites.
  - This was a workaround to the relatively small presence of longitudinal data.
- Final working data set
  - Observations for 1308 unique participants
  - 608 had complete data (at least one result from each test & no missing covariates).

Response Patterns of Positive Diagnosis with (Microscopy, PCR, RDT)

Response	Count	Response	Count
(1,1,1)	41	(1,m,1)	66
(1,1,0)	22	(1,m,0)	21
(1,0,1)	11	(0,m,1)	153
(0,1,1)	78	(0,m,0)	312
(1,0,0)	23	(m,1,1)	5
(0,1,0)	80	(m,1,0)	3
(0,0,1)	74	(m,0,1)	4
(0,0,0)	280	(m,0,0)	6
(1,1,m)	0	(1,m,m)	1
(1,0,m)	0	(0,m,m)	1
(0,1,m)	1	(m,m,1)	47
(0,0,m)	1	(m,m,0)	80

The following responses were not observed: (1,1,m), (1,0,m), (m,1,m), (m,0,m)

Comparing Basic LCA models (no covariates)

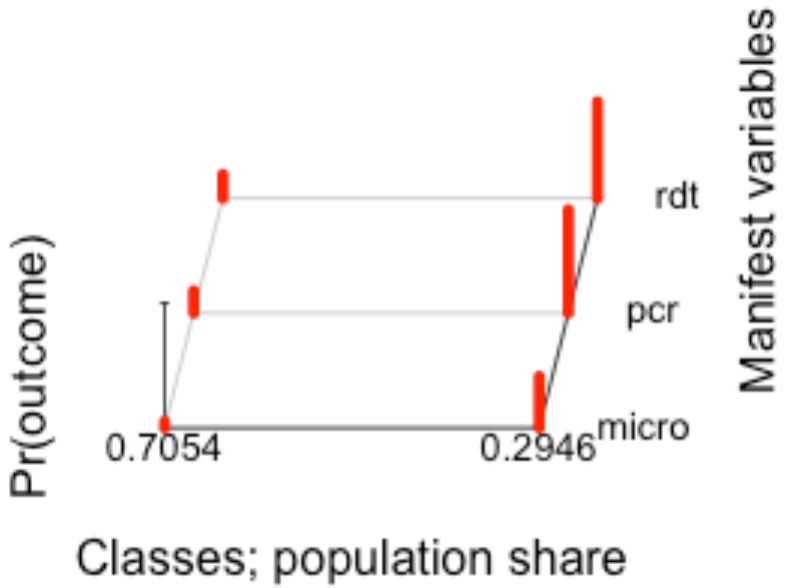
# classes	# cells	df**	Model parameters to be estimated	Residual df***	AIC*	BIC*	G^2 (LRT statistic)
1	8	7	3	4	3568.4	3584.0	112.4
2	8	7	7	0	3416.1	3452.3	5.45

* Lower is better

** # degrees of freedom =
# cells in contingency table - 1

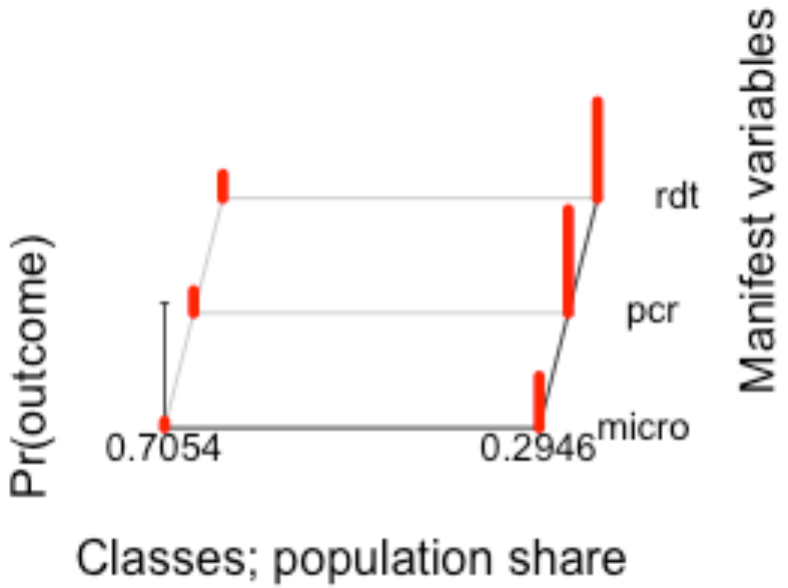
*** Residual degrees must be nonnegative in order for the model to be identifiable

Basic LCA model with 2 classes: Probabilities of positive diagnosis



Diagnostic method	Latent class 1: Lower prevalence rates	Latent class 2: Higher prevalence rates
Microscopy	0.05	0.41
PCR	0.18	0.83
RDT	0.19	0.78

Basic LCA model with 2 classes: Probabilities of positive diagnosis



Diagnostic method	Latent class 1: Lower rates of (+) diagnoses	Latent class 2: Higher rates of (+) diagnoses
Microscopy	0.05	0.41
PCR	0.18	0.83
RDT	0.19	0.78

Comparing LCA models with the covariate *firstpreg*

# classes	# cells	df	Model parameters to be estimated	Residual df	AIC	BIC	Negative max log-likelihood
1	8	7	3 *	4	3562.8	3578.3	112.2
2	8	7	8	-1 **			

* The LCA model with one class is not affected by the addition of covariates. (The slight differences from the previous one-class model in the goodness-of-fit measures are due to the discarding of a few observations missing a value for *firstpreg*.)

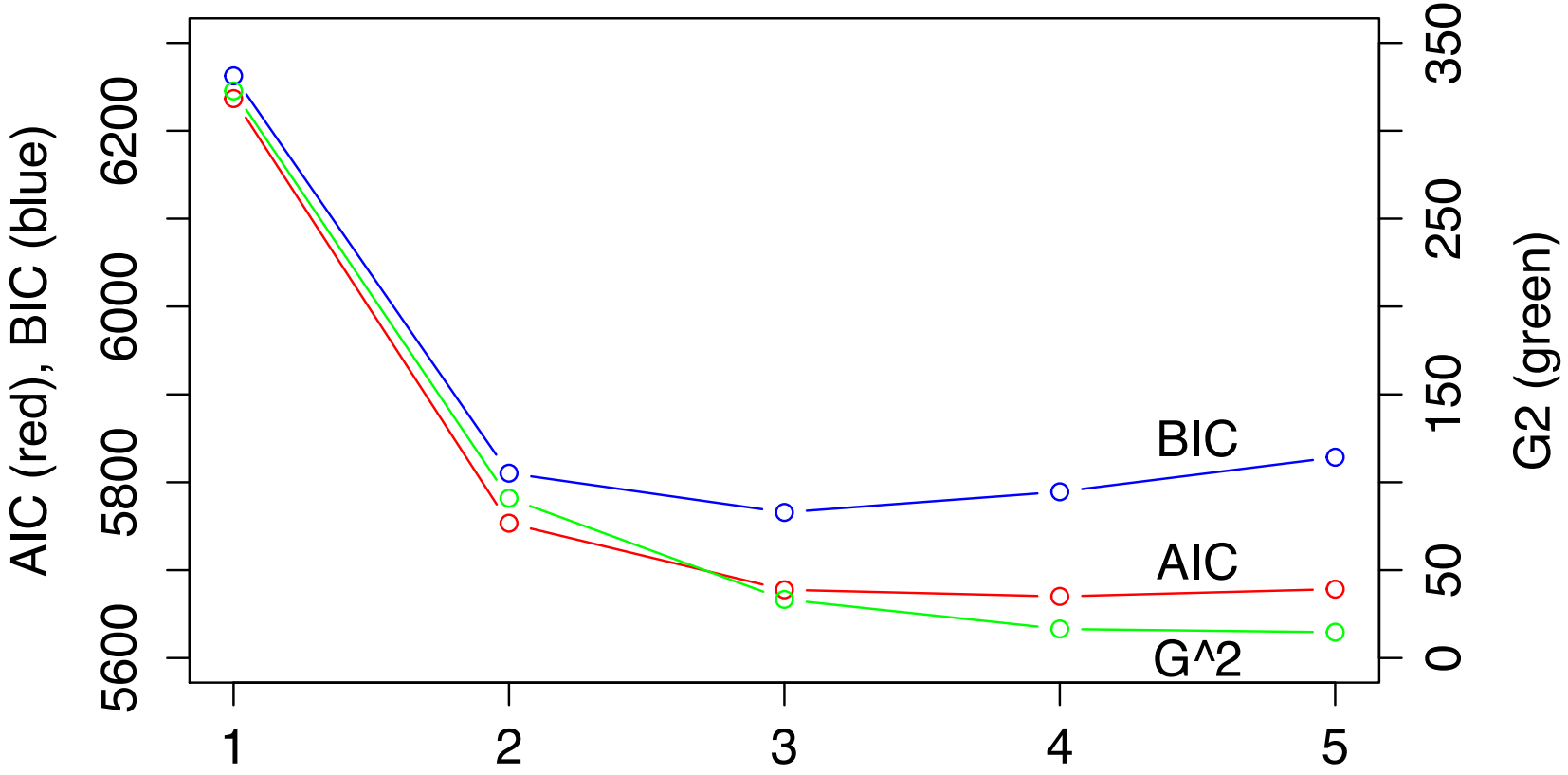
** The model requires the estimation of an additional parameter (the effect size of the covariate on the latent variable). This makes non-identifiable any model with more than one latent class.

Comparing LCA models treating *firstpreg* and *teen* as manifest vars.

# classes	# cells	df	Model parameters to be estimated	Residual df	AIC	BIC	Negative max log-likelihood
1	32	31	5	28	6236.6	6262.5	322.7
2	32	31	11	20	5753.4	5810.4	90.9
3	32	31	17	14	5677.7	**5765.7**	33.3
4	32	31	23	8	**5670.0**	5789.2	16.5
5*	32	31	29	2	5678.4	5828.5	14.7

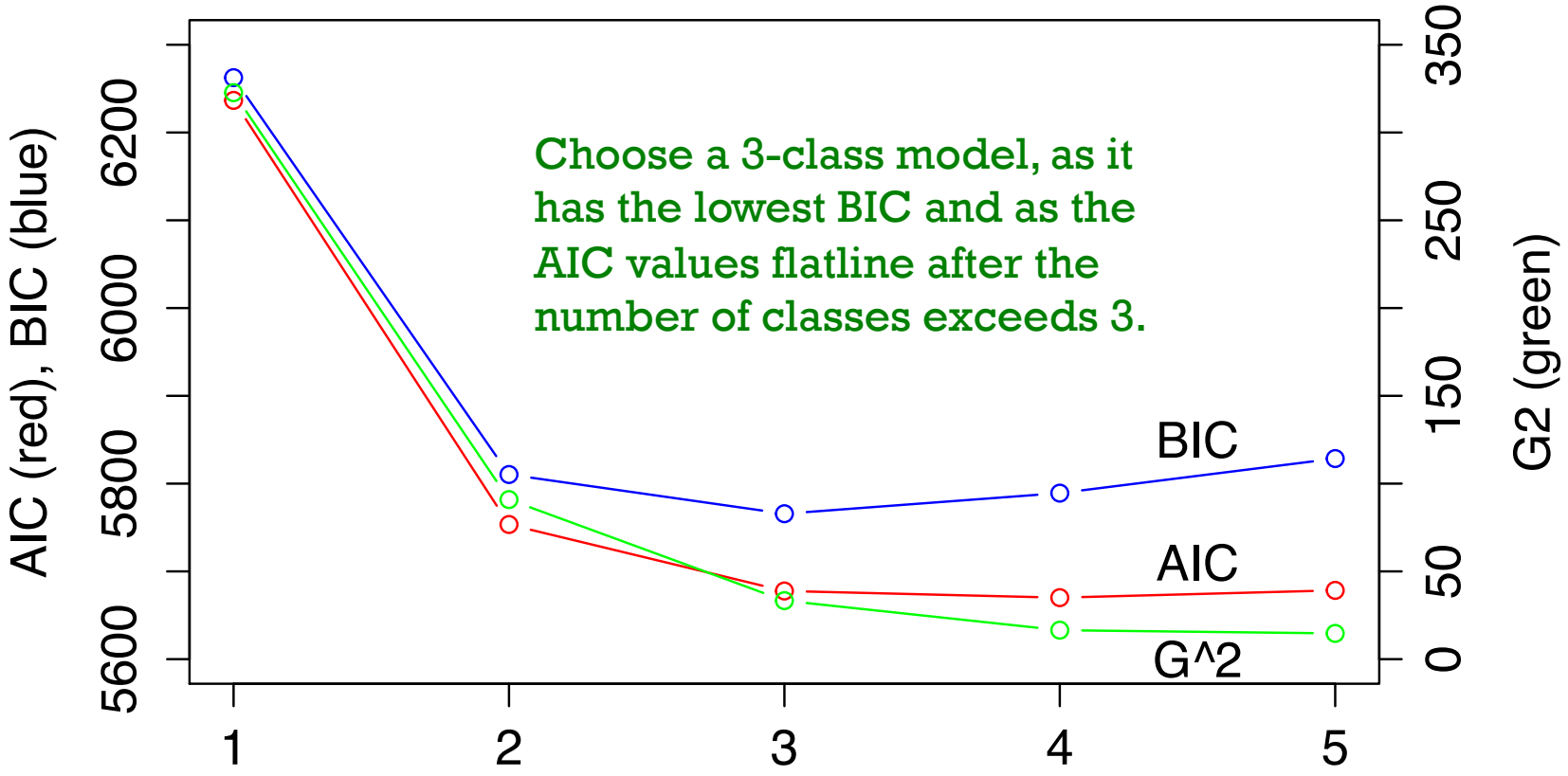
* The MLE for the 5 class model was not found by the software poLCA.

Comparing LCA models treating *firstpreg* and *teen* as manifest vars.



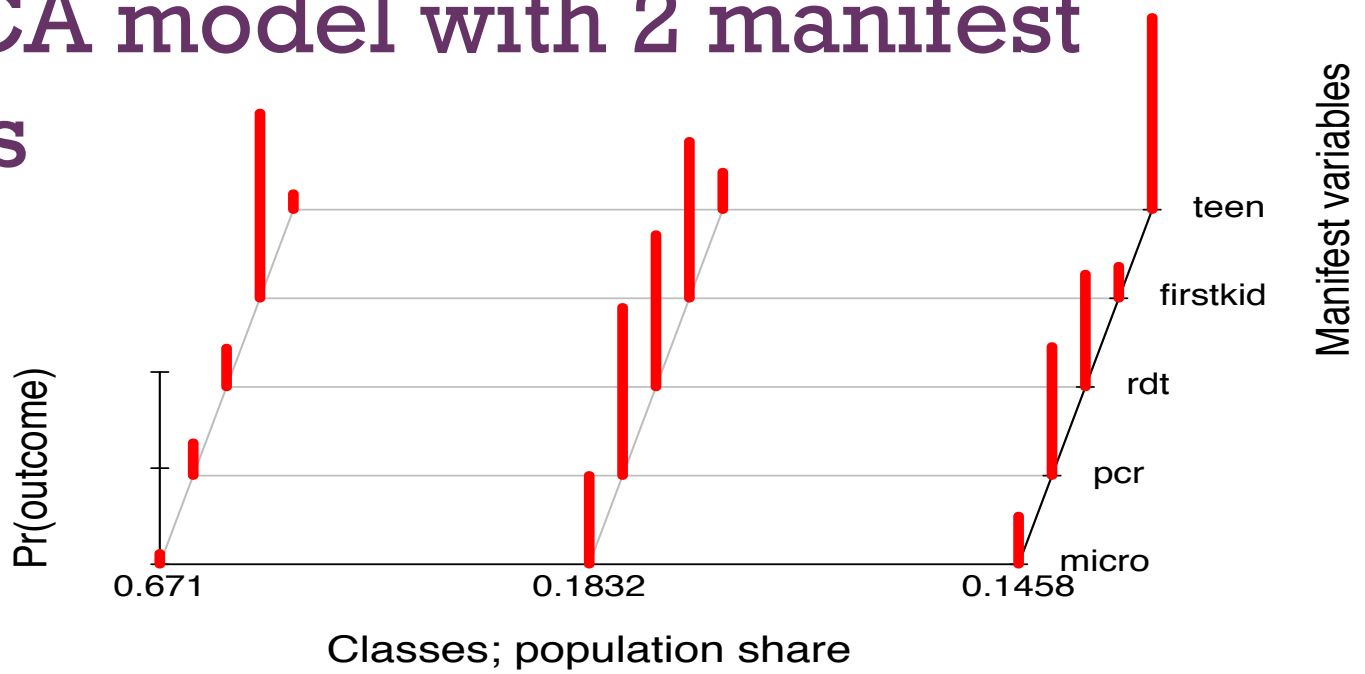
The MLE for the 5 class model was not found by the software poLCA.

Comparing LCA models treating *firstpreg* and *teen* as manifest vars.



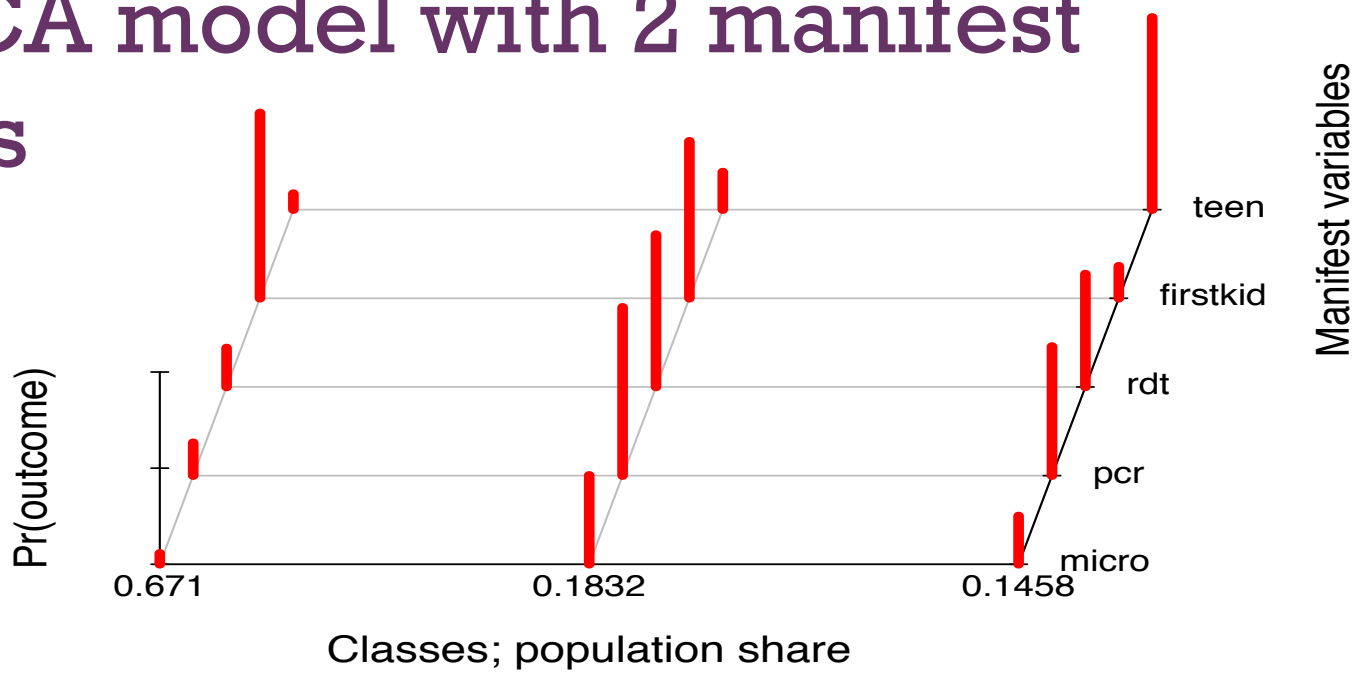
The MLE for the 5 class model was not found by the software poLCA.

3-class LCA model with 2 manifest covariates



Manifest variable	Latent class 1: First pregnancies	Latent class 2: Highest prevalences	Latent class 3: Teenage mothers
Microscopy	0.06	0.46	0.25
PCR	0.17	0.88	0.67
RDT	0.20	0.80	0.59
firstpreg	0.96	0.82	0.17
teen	0.08	0.20	1.00

3-class LCA model with 2 manifest covariates



Manifest variable	Latent class 1: First pregnancies	Latent class 2: Highest rate of (+) diagnoses	Latent class 3: Teenage mothers
Microscopy	0.06	0.46	0.25
PCR	0.17	0.88	0.67
RDT	0.20	0.80	0.59
firstpreg	0.96	0.82	0.17
teen	0.08	0.20	1.00

Summary

- Basic LCA models
  - PCR and RDT methods have concordance within 5% of each other, even when probabilities of positive diagnosis are below 20%.
  - Weak concordance between microscopy and the other two methods.
- LCA models with covariates are non-identifiable
- LCA models treating covariates as manifest
  - Concordance between PCR and RDT within 8%.
  - This model provides weak, if any, evidence for causal inference.
- Suggestions for future work
  - Setting parameter restrictions to ensure identifiability of models with covariates.
  - Treating covariates as continuous
  - Treating manifest variables as continuous → Latent trait analysis
  - Include an indicator for household income as a covariate
    - Teenage mother with multiple children ←→ Poverty

References and Acknowledgements

- Thanks to Dr. Steven Meshnick and Dr. Steve Taylor for the opportunity to work on the project data set.

- Some references:

  - Collins LM and Lanza ST. (2010) *Latent Class and Latent Transition Analysis*. Wiley.
  - Linzer, DA and Lewis J. (2011) "poLCA: an R Package for Polytomous Variable Latent Class Analysis." *Journal of Statistical Software*. 42(10): 1-29. http://www.jstatsoft.org/v42/i10