Assessing Concordance among Diagnostic Tests for Malaria using Latent Class Analysis[1]

Introduction

Methods of detecting the malarial parasite *Plasmodium falciparum* exist, include (1) microscopy, (2) microscopy following PCR, and (3) rapid diagnostic tests (RDT). None of these, however, is considered to be a gold standard for all situations. The absence of a universal gold standard malarial diagnostic test prevents the direct evaluation of sensitivity and specificity of any of these methods. Methods involving microscopy have lower sensitivity when parasites are at low densities. Furthermore, the time required to collect and process blood samples can make microscopy methods costly and time-consuming for screening populations with low prevalence of malaria. (There have been recent proof-of-concept efforts to harness crowdsourcing for diagnosis of blood samples[2,3], but it remains to be seen how well crowdsourcing could mitigate the time lag problem.) RDT's which target specific antigens related to *P. falciparum* give results that can be read on-site, and thereby have the potential to reduce the speed and cost of screening for malaria. Their sensitivities and specificities, however, are even less well established than those of methods involving microscopy.

We can resort to more indirect methods, however, by assessing concordance or "rater agreement" between different diagnostic tests. There are at least a few options for assessing rater agreement. One of the more simple to calculate and widely used index for rater agreement is the Cohen's kappa statistic, which is typically used for agreement between two raters. The kappa statistic, however, has its share of detractors, as can produce different results based only on changes in prevalence.[4]

Latent class analysis (LCA) is an exploratory statistical approach that has been used to assess rater agreement for diagnostic tests in the absence of a gold standard. The aim of LCA is to partition the sample population into a finite (and typically small) number of latent classes that are distinct from each other with respect to diagnostic test results. These latent classes are not defined beforehand. Rather, LCA works on the premise that that the unobserved membership in a latent class can be indirectly predicted via observed 'manifest' variables. LCA assumes conditional independence of these manifest variables; that is, it is assumed that within each latent class the manifest variables are independent of each other. Some models also include covariates as predictors of latent class membership. After the classes are separated, it remains up to the researcher to describe how the classes differ from each other; in this sense it is a qualitative assessment. It is hoped that LCA will shed light on the strength of agreement between malarial diagnostic tests, whether the prevalence of malaria is low or high.

Methods

Data cleaning and preparation

Four data sets (d2, d6a, d6b, d7) were obtained from iLiNS (International Lipid-Based Nutrient Supplements) DYAD Project through the help of Dr. Steve Meshnick of UNC and Dr. Steve Taylor of Duke University. The data set records were merged according to Participant ID. Some data cleaning was carried out to clear out redundant or ambiguous records, such as (1) any record which duplicated another record with respect to participant, categorical time of visit, and test result, (2) records with the same Participant ID but with conflicting test results. The statistical programming language R 2.15.2 was used to analyze, extend, and merge the data sets.

LCA requires explanatory variables to be categorical, and so to accommodate this the results of each of the three malarial diagnostic tests were dichotomized naïvely into zero vs non-

zero values, with the exception of defined error/missing codes ('99' for RDT's, and '999' for microscopic assays).

Timepoints were also divided into two categories, second trimester versus third trimester. Observations with "mEli" listed for the variable "NumberVisit" were considered to have occurred during the 2nd trimester, while all others ("m26c", "m32c","m36c","c36c","c00h") were considered to have occurred during the 3rd trimester.

For individuals with two results from different trimesters for the same kind of test, a "once positive, always positive" approach was taken to simplify the data. That is, an individual was considered to have a positive result for a particular test if any iteration of that particular test showed a positive result. An argument in favor of this approach is that not all blood samples from the same individual with malaria may show parasites.

Observations for individuals who had two contradictory results for the same kind of test during the same trimester were discarded from the data set to be analyzed. Often these conflicts occurred because participant ID's were shared by two people (having different ages and birthdays).

Covariates considered were age and gravidity. Age was dichomotized by the variable *teen* indicating whether age was under 20 years old. Gravidity was dichotomized by the variable *firstpreg* indicating whether the woman had any previous pregnancies.

Modeling

The R package poLCA[5] was used to generate LCA models without covariates. poLCA uses a maximum likelihood estimation approach, with a mixture of the EM and Newton-Raphson algorithms.

Three categories of models were considered: (1) LCA models with no covariates, (2) LCA models using covariates (see), and (3) LCA models treating covariates as manifest

variables. Only models of types (1) and (2) were considered in the statistical analysis plan, but after realization that models of type (2) were non-identifiable for our purposes, models of type (3) were considered. (2) and (3) differ in that (2) treats covariates as predictors of the unobserved latest class membership, while (3) assumes it is latent class membership which affects the covariates.

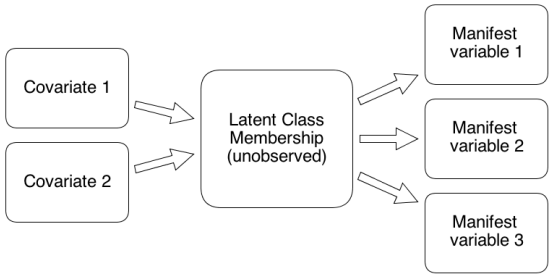

Figure 1: Schematic representation of a latent class model with covariates predicting latent class membership.

While comparing models that differ only in the number of latent classes, the number of classes can be chosen using an informal holistic assessment of AIC, BIC, and the second-order changes of G^2 (likelihood ratio test statistic) and the maximum log likelihood as the number of latent classes increases. (There is some disagreement on the use of the likelihood ratio test (LRT) to determine the number of classes, since it does not always have an approximate chi-square distribution for testing the number of latent classes.[6] There does not appear to exist another test statistic that is widely accepted and available in a software package. The Lo-Mendell-Rubin adjusted LRT was proposed in 2001, but its proof was contested in 2003, and it does not appear to yet be available in R or SAS.)

Results

After merging the data tables by *Participant* and *NumberVisit*, the final working data set had 1316 unique participant ID's, of which 608 had at least one result for each of the three diagnostic tests and no missing covariates. Table 1 shows the patterns of positive diagnosis for malaria by different test and the number of individuals for each response pattern.

Table 1: Frequencies of patterns of positive summary diagnoses for malariafrom 3 types of diagnostic tests for malaria (microscopy, PCR, RDT). Key: 0=negative diagnosis; 1=positive diagnosis; m=missing value

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,m,m)	1
(1,1,m)	0	(0,m,m)	1
(1,0,m)	0	(m,1,m)	0
(0,1,m)	1	(m,0,m)	0
(0,0,m)	1	(m,m,1)	47
		(m,m,0)	80

LCA models with no covariates

With 3 dichotomous tests, only LCA models with either 1 or 2 can be specified. A model with 3 classes cannot be estimated, since the number of parameters to be estimated exceeds the number of cells in the 2x2x2 contingency table. (See Table 2.)

Table 2: A worked example of how to determine whether a latent class model is identifiable

# latent classes	# different results for			# diff. combos of results	# parameters in LC model to be estimated	Is the LC model identifiable?
	micro-scopy	PCR	RDT			
R	K_1	K_2	K_3	$\prod K_j$	$R\sum(K_j-1)+(R-1)$	$\prod K_j - 1 \geq R\sum(K_j-1)+(R-1)$
1	2	2	2	8	3	yes
2	2	2	2	8	7	yes
3	2	2	2	8	11	no

For models without parameter restriction, models with 1 and 2 classes were generated. A model with 3 classes was also specified, but did not converge as it had a negative number of degrees of freedom.

Goodness-of-fit statistics for the LCA models are in Table 3.

Table 3: Goodness of fit measurements for selecting the number of latent classes for the model without covariates.

# latent classes	df = # cells in contingency table - 1	parameters to estimate	Residual df	AIC	BIC	G^2
1	7	3	4	3568.4	3584	112.4
2	7	7	0	**3416.1**	**3452.3**	5.45

Only models with a non-negative residual degrees of freedom are identifiable; otherwise a unique solution cannot be estimated. Lower scores for AIC and BIC generally indicate a better model fit; accordingly, the model with two classes was selected.

The probabilities of positive test diagnoses within each class of a two-class model in seen in Table 4. The concordance between PCR and RDT is within 5 percentage points of each other, whether the probabilities of positive diagnoses are high or low (below 20%).

Table 4: Probabilities of positive test diagnoses for a two-class LCA model without covariates

	Latent class 1: Lower rates of positive diagnoses	Latent class 2: Higher rates of positive diagnoses
Latent class prevalences	0.71	0.29
Diagnostic method		
Microscopy	0.05	0.41
PCR	**0.18**	**0.83**
RDT	**0.19**	**0.78**

LCA models with covariates

The addition of a single covariate to the model raises the number of parameters to be estimated by one. For the two-class model this results in a non-identifiable model, as indicated by a negative residual degrees of freedom. (See Table 5.) That is, estimates of the model will not be consistent. The two-class LCA model with covariates would be feasible only if we had 4 or more diagnostic tests altogether.

Table 5: Goodness of fit measurements for selecting the number of latent classes for the model with dichotomous covariate for previous pregnancy.

# latent classes	df = # cells in contingency table - 1	parameters to estimate	Residual df	AIC	BIC	G^2
1	7	4	3	3562.8	3578.3	112.2
2	7	8	-1	-	-	-

Note that the one-class LCA model with *firstpreg* is not affected by the addition of covariates, since adding covariates does not add any power to predicting latent class membership. (The slight differences from the previous one-class model in Table 3 are due to the discarding by poLCA of a few observations missing a value for *firstpreg*.)

Similar results (not shown) were seen for LCA models with *teen* as a covariate.

LCA models treating covariates as manifest variables

Assessing the relationship between the covariates *firstpreg* and *teen* with the concordances of diagnostic test results was not possible by treating the covariates as predictors of latent class membership. If we instead treat the covariates as manifest variables, we can obtain identifiable models at the expense of weakening the interpretability of the model (such as with regard to causal inference). LCA models in which covariates appear as manifest variables regard the covariates as conditionally independent of the diagnostic tests. Despite this drawback, we are still able to assess concordances among the diagnostic tests within each latent class.

Among these models, a 3-class model was chosen, as it has the lowest BIC and as the AIC values flatlined after the number of classes exceeded three. (See Table 6 and Figure 2.)

Table 6: Goodness of fit measurements for selecting the number of latent classes for the model treating dichotomous covariates for previous pregnancy and age under 20 as manifest variables.

# latent classes	df = # cells in contingency table - 1	parameters to estimate	Residual df	AIC	BIC	G^2
1	31	5	28	6236.6	6262.5	322.7

2	31	11	20	5753.4	5810.4	90.9
3	31	17	14	5677.7	**5765.7**	33.3
4	31	23	8	**5670.0**	5789.2	16.5
5*	31	29	2	5678.4	5828.5	14.7

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

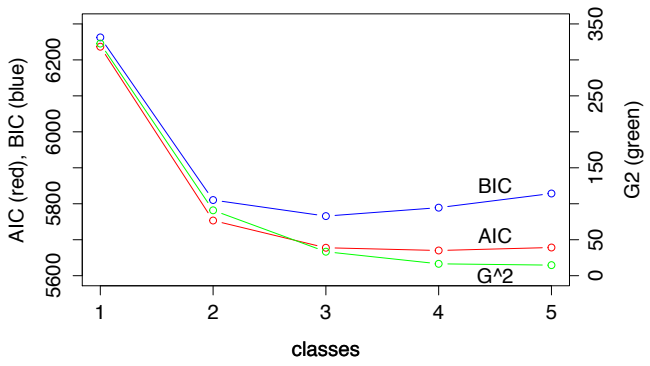

Figure 2: Goodness of fit measures for selecting the number of latent classes for the model treating dichotomous covariates for previous pregnancy and age under 20 as manifest variables.

Within each latent class in the three-class model, the concordances between PCR and RDT were within 8 percentage points of each other. (See Table 7.) This concordance appears to be strongest among mothers at least 20 years of age who were pregnant for the first time who live in areas with low prevalence of malaria. The concordance was still noticeable among teenage women who had had a previous pregnancy prior to participation in the study.

Table 7: Probabilities of positive test diagnoses for a two-class LCA model treating dichotomous covariates for previous pregnancy and age under 20 as manifest variables.

	Latent class 1: First pregnancies	Latent class 2: Highest positive diagnoses	Latent class 3: Teenage mothers
Latent class prevalences	0.67	0.18	0.15
Diagnostic method			
Microscopy	0.06	0.46	0.25
PCR	**0.17**	**0.88**	**0.67**
RDT	**0.20**	**0.80**	**0.59**
"Manifest covariates"			
firstpreg	0.96	0.82	0.17
teen	0.08	0.20	1.00

Discussion

In general, the concordance between PCR and RDT tests appears to be good, even among populations with low prevalence of malaria. This suggests the possibility that RDT may be a cost-effective substitute for PCR for diagnosing malaria.

We should see these results, however, not as definitive but as suggestive, given the assumptions and limitations of our models: (1) Given that we have no gold standard, we must be careful not to confuse observed test results with actual disease status. (2) We argue that a "once positive, always positive" assumption is reasonable since malaria has no apparent cure and since it is possible that not all samples taken from a individual with malaria may carry the parasite. That said, this assumption may lead to overestimates of prevalence rates, particularly if more test readings are taken. (3) LCA models assume conditional independence between the malarial diagnostic tests; that is, within any latent class the tests are assumed to be independent of each other. Violation of this assumption may result in exaggerated conclusions about the strengths of concordances.

Missing data would have proved a challenge to LCA modeling, for if we had tried to look at the data longitudinally, we would have had to deal with a great deal of missing data. The use of the "once positive, always positive" assumption allowed us to collapse the data longitudinally and in so doing allowed us to avoid having such a large proportion of the data missing. Once test results were collapsed longitudinally, the remaining missing data did not pose a major obstacle for modeling, as poLCA can handle missing values for manifest variables. poLCA, however, discards observations with missing values for covariates, but there were only a couple of observations that were discarded in our analysis.

Further research could include the following: (1) considering other dichomtomizing cut-points for the manifest variables or the covariates, (2) setting parameter restrictions to make latent class models with covariates identifiable, (3) use of Bayesian methods to estimate standard errors. (Estimation of standard errors is difficult using maximum likelihood methods.)

---

[1] Statistical consultant: Noory Kim. Researcher client: Steven Meshnick MD PhD, Department of Epidemiology, UNC. meshnick@email.unc.edu

[2] Mavandadi S, et al. "Distributed Medical Image Analysis and Diagnosis through Crowd-Sourced Games: A Malaria Case Study." *PLoS ONE* 2012, 7(5):e37245.

[3] Luengo-Oroz MA, Arranz A, and Frean J. "Crowdsourcing Malaria Parasite Quantification: An Online Game for Analyzing Images of Infected Thick Blood Smears." *J Med Internet Res* 2012, 14(6):e167. www.jmir.org/2012/6/e167. Game website: www.malariaspot.org.

[4] Uebersax J. "Kappa Coefficients." www.john-uebersax.com/stat/kappa.htm, updated 1 October 2009, accessed 1 April 2013.

[5] Linzer DA and Lewis J. "poLCA: Polytomous Variable Latent Class Analysis." R package version 1.4. 2013. http://dlinzer.github.com/poLCA.

[6] Nylund KL, Asparouhov T, and Muthén BO. "Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study" *Structural Equation Modeling* 2007, 14:535–569.