LDA 140.665 Final Exam
Due: 03/21/2005

Question (a)

Model for $E(y_{ij})$: $\log it\mu_{ij} = \log(\frac{\mu_{ij}}{1-\mu_{ij}}) = \beta_0 + \beta_1 time_{ij} + \beta_2 age_i + \beta_3 sex_i + \beta_4 vitA_i$

where $E(y_{ij}) = \mu_{ij}$

Model for variance: $\text{var}(y_{ij}) = \mu_{ij}(1-\mu_{ij})$

Question (b): under the model for $E[y_{ij}]$ in (a)

(i) The *probability* that a female child age 4 who does not have vitamin A deficiency will *NOT* have a respiratory infection at the final visit

Given: $time_{ij}$=15 (last visit), age=4, sex=1, vitA=0 (not deficient)

$\text{logit}\,\mu_{ij} = \beta_0 + \beta_1 * 15 + \beta_2 * 4 + \beta_3 * 1 + \beta_4 * 0 = \beta_0 + 15\beta_1 + 4\beta_2 + \beta_3$

→ **Prob (*not* having a respiratory infection)** = 1-P(having a respiratory infection)

$$= 1 - \frac{\exp(\beta_0 + 15\beta_1 + 4\beta_2 + \beta_3)}{1 + \exp(\beta_0 + 15\beta_1 + 4\beta_2 + \beta_3)} = \frac{1}{1 + \exp(\beta_0 + 15\beta_1 + 4\beta_2 + \beta_3)}$$

(ii) The odds that a male child of age 3 with vitamin deficiency will have a respiratory infection at the initial visit

Given: $time_{ij}$=0 (initial visit), age=3, sex=0, vitA=1 (deficient)

$\text{logit}\,\mu_{ij} = \beta_0 + \beta_1 * 0 + \beta_2 * 3 + \beta_3 * 0 + \beta_4 * 1 = \beta_0 + 3\beta_2 + \beta_4$

→ **The odds (having a respiratory infection)** =$\exp(\beta_0 + 3\beta_2 + \beta_4)$

(iii) β_4**>0**

[i.e. β_4 is the regression coefficient for vitA (1=deficient; 0=not deficient)]

Question (c): without taking correlation into account

Table 1: Results from logistic regression (assume independent responses)
(STATA: logit RI time age female vitA)

RI	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	.0168947	.0111999	1.51	0.131	-.0050567	.0388462
age	-.0715987	.0297275	-2.41	0.016	-.1298636	-.0133338
female	-.5576829	.114908	-4.85	0.000	-.7828984	-.3324674
vitA	**.2869989**	**.1175161**	**2.44**	**0.015**	**.0566716**	**.5173263**
_cons	-.5575417	.162073	-3.44	0.001	-.8751989	-.2398845

Null Hypothesis H_0: β_4**=0** (i.e. vitamin A deficiency is not associated with respiratory infection)
Test statistics:

Wald test: $z = \frac{\hat{\beta}}{se(\hat{\beta})} = \frac{0.2869989}{0.1175161} = 2.44$, p-value=0.015 (<0.05)

[$z^2 = \chi^2_{(1)}$ =5.96, p-value= Prob($\chi^2_{(1)}$ >5.96)= 0.0146<0.05)]

Or we can also use *likelihood ratio test:* χ^2 =-2[log likelihood (null)-log likelihood (extended)] = -2[(-894.23474)-(-891.27072)] = 5.93;
P-value= 0.0149
Conclusions: there is evidence suggesting that vitamin A deficiency is associated with the risk of having respiratory infection after controlling for age, gender and time (follow-up period). The odds of having respiratory infection among children with vitamin A deficiency is 1.33 (=exp (0.287)) times the odds of having respiratory infection among children of similar age, gender and follow-up time without vitamin A deficiency (OR=1.33; 95% CI: 1.06 to 1.68).

Question (d)

If we do not take the correlation between repeated measurements on the same subject into account, the estimates of regression coefficients are inefficient. That is, they are more variable (less precision) than they could be if the correlation was taken into appropriate account. More importantly, the standard errors and the inferences (such as confidence intervals, t test, Wald test…etc) are incorrect. Therefore, the inference

about the regression coefficients will be wrong.

Question (e): taking correlation among repeated measurements into account
Model for the mean:

$$\log it\mu_{ij} = \log(\frac{\mu_{ij}}{1-\mu_{ij}}) = \beta_0 + \beta_1 time_{ij} + \beta_2 age_i + \beta_3 sex_i + \beta_4 vitA$$

$Corr(Y_{ij}, Y_{ik}) \neq 0$
We may extend the model (a) to allow for by various possible correlation structures, including *unstructured, exchangeable (uniform),* and *AR(1),* and look for the consistency between these models.

Table 2: Comparisons of regression coefficients of Vitamin A deficiency obtained under different assumptions of within-subject correlation structure (using GEE method)

	Vitamin A deficiency (yes versus no)				
	Coefficient	SE	P-value	Robust SE	P-value (with robust variance estimation)
Unstructured	0.284	0.2189	0.195	.2206008	0.198
Exchangeable	0.276	0.2207	0.211	.2232926	0.217
AR(1)	0.286	0.1804	0.113	.2242829	0.202

STATA command:

```
xtgee RI time age female vitA, nolog f(bin) link(logit) corr(uns) robust
xtgee RI time age female vitA, nolog f(bin) link(logit) corr(exc) robust
xtgee RI time age female vitA, nolog f(bin) link(logit) corr(ar1) robust
```

In general, the standard error with robust variance estimation approach is close to the standard errors without robust estimation, except for the AR(1) model. The coefficients of vitamin A deficiency estimated from all these correlation models are very close to each other. Based on robust variance estimation, we got the same inference about the coefficient of vitamin A deficiency no matter which correlation model we assume. That is, the inference suggested that vitamin A deficiency does not significantly change the risk of respiratory infection.

Question (f)

The model from question (e) that has the fewest assumptions is the *unstructured model.* Assuming that unstructured model is correct, we test whether the mean pattern of respiratory

is associated with the presence or absence of vitamin A deficiency. The results of the regression coefficients (with robust variance estimation) are shown in Table 3:

Table 3. GEE results with unstructured correlation matrix (with robust variance estimation)

RI	Semi-robust Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	.017177	.0082279	2.09	0.037	.0010506	.0333034
age	-.0771982	.0542186	-1.42	0.154	-.1834647	.0290683
female	-.5339991	.2151933	-2.48	0.013	-.9557703	-.112228
vitA	**.2840136**	**.2206008**	**1.29**	**0.198**	**-.148356**	**.7163833**
_cons	-.5469366	.2742036	-1.99	0.046	-1.084366	-.0095073

H_0: $\beta_4=0$

Wald test: $z=\frac{\hat{\beta}}{se(\hat{\beta})}=\frac{0.284}{0.2206}=1.29$; p-value=0.198 (>0.05)

[$z^2=\chi^2_{(1)}=1.66$, p-value= Prob($\chi^2_{(1)}>1.66$)= 0.1979 >0.05)]

Conclusion: There isn't sufficient evidence (p-value >0.05) supporting that vitamin A deficiency is associated with the risk of developing respiratory infection after controlling for age, gender and time (follow-up period). This result based on unstructured correlation model does NOT agree with that in part (C), which found a significant association between vitamin A deficiency and respiratory infection under the assumption of independence (i.e. no correlation between repeated measurements on the same subject). The reason for the different results is that the standard error under the assumption of independence (i.e. S.E=0.1175) was much smaller than the standard error with correlation taken into account (i.e. SE=0.2206). Therefore, we got the wrong inference about the regression coefficient while we assume the outcomes are independent. The inference based on the unstructured correlation matrix should be the correct one.

Question (g)

The estimated within-subject correlation matrix from unstructured model is provided below (Table 4). Based on the matrix, it seems to me that the correlation matrix is close to an exchangeable (uniform) correlation structure. We compare this unstructured matrix with the correlation structure estimated under the exchangeable model (Table 5). They look quite similar to each other. Therefore, I think that a *exchangeable* (*uniform) correlation model* would be an appropriate and simpler model to the data in this case.

Table 4: Estimated within-subject correlation matrix (Unstructured)

	c1	c2	c3	c4	c5	c6
r1	1.0000					
r2	0.5623	1.0000				
r3	0.4606	0.5757	1.0000			
r4	0.4240	0.5629	0.5587	1.0000		
r5	0.5035	0.5251	0.4250	0.4342	1.0000	
r6	0.4480	0.5636	0.5148	0.5189	0.5097	1.0000

Table 5: Estimated within-subject correlation matrix (Exchangeable Model)

	c1	c2	c3	c4	c5	c6
r1	1.0000					
r2	0.5060	1.0000				
r3	0.5060	0.5060	1.0000			
r4	0.5060	0.5060	0.5060	1.0000		
r5	0.5060	0.5060	0.5060	0.5060	1.0000	
r6	0.5060	0.5060	0.5060	0.5060	0.5060	1.0000

We now estimate the coefficients based on exchangeable (uniform) model. The results (with robust variance estimation) are in Table 6

Table 6: GEE results with exchangeable correlation matrix (with robust variance estiamation)

	Semi-robust					
RI	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	**.0168792**	**.0082807**	**2.04**	**0.042**	**.0006493**	**.033109**
age	-.0744012	.0546842	-1.36	0.174	-.1815803	.032778
female	**-.5550523**	**.2168697**	**-2.56**	**0.010**	**-.9801092**	**-.1299955**
vitA	.2757902	.2232926	1.24	0.217	-.1618552	.7134357

```
       _cons |  -.5443887   .2759358    -1.97   0.049    -1.085213   -.0035644
------------------------------------------------------------------------------
```

(i) Because the result, the p-value for the regression coefficient of the time variable is 0.042, which is less than 0.05. Therefore, there is evidence suggesting that the probability of respiratory infection changed over the 15 month study period (estimate of coefficient= 0.017, 95% CI=0.0006 to 0.0331)

(ii) Yes, there is sufficient evidence showing that gender will affect the risk of respiratory infection in this population of children (estimate of coefficient= -0.555, 95% CI=-0.98 to -0.13, p-value= 0.01). It is worthwhile to take gender into account.

Question (h)

Based on model in (g) (i.e. exchangeable model), we know the log odds that a female child of age 7 with vitamin A deficiency has respiratory infection at the initial visit:

$$\log it\mu_{ij} = \log(\frac{\mu_{ij}}{1-\mu_{ij}}) = (-0.544) + (0.017)time_{ij} + (-0.074)age_i + (-0.555)sex_i + (0.276)vitA$$

$$= (-0.544) + 7*(-0.074) + (-0.555) + 0.276 = -1.344$$

The probability that a female child of age 7 with vitamin A deficiency has a respiratory infection at initial visit

$$= \frac{\exp(-1.344)}{1+\exp(-1.344)} = \mathbf{20.7\%}$$

Summary

In this study, we are interested in determining the effect of vitamin A deficiency on developing respiratory infection in preschool children. The data of 250 children participating in a study in Indonesia for a period of 15 months was used.

Marginal logistic regression model with the exchangeable (uniform) correlation structure was used to evaluate the association between vitamin A deficiency and respiratory infection. We assume that the correlation between repeated measurements of outcomes within the same subject followed the exchangeable (uniform) correlation model. This is because the estimated uniform correlation matrix from the uniform

model is similar to that from the unstructured model. Therefore, uniform correlation model could be an appropriate but simpler model for correlation. Logistic model is chosen because of the outcome (i.e. respiratory infection) is binary (yes or no). The covariates included in the model are vitamin A deficiency status, time (i.e. follow-up study period), age (at entry), and gender because we worry that the time, age at entry and gender may confound the effect of vitamin A deficiency on the outcome. The model assumes that the log odds of respiratory infection follows a linear combination of intercept and additive terms for time, age, gender and vitamin A.

Based on the results from the marginal logistic model with exchangeable correlation (with robust variance estimation), the population-average odds of infection among vitamin A deficient children is about 1.32 [1.32=exp (0.276), 95% CI= 0.85 to 2.04] times the odds of infection among otherwise similar (i.e. of the same age, gender and follow-up time) children population without vitamin A deficiency (Table 6). This result suggests that vitamin A deficiency did not significantly change the risk of having respiratory infection after controlling for all other covariates (age at entry, gender and time) in the model. However, the risk of respiratory infection appeared to significantly change over the 15 month study period and to be affected by the gender.

We also fit the logistic regression with random intercept and the transition model to evaluate the effect of vitamin A deficiency on respiratory infection. Based on the random intercept model, the odds of infection for a child with random intercept Ui when he/she is vitamin A deficient is 1.93 (95% CI= 0.84 to 4.45) times the odds of infection when the same child (with random intercept Ui) is not vitamin A deficient. The regression coefficient from marginal model is much smaller than that from the random intercept model. This suggests that the variance of random intercept is large, and there is substantial heterogeneity across children in the propensity of respiratory infection not attributable to these covariates. Based on the transition model, we noticed that the previous respiratory infection is a significant predictor of current respiratory infection (coefficient estimate=2.55, 95% CI= 2.14 to 2.96). The odds of infection for a child with outcome at the previous visit y_{ij-1} when he/she is vitamin A deficient is 1.17 (95% CI= 0.85 to 1.60) times the odds of infection when the same child is not vitamin A deficient. The estimates of the effect of vitamin A deficiency on respiratory infection based on marginal model, random intercept model and transition model are summarized in Table 7.

In conclusion, there is no sufficient evidence suggesting that the vitamin A deficiency is associated with the risk of respiratory infection after controlling for the effect of age

at entry, time, and gender.

Table 7. Estimate of coefficient and odds ratio (OR) for vitamin A deficiency under different models.

	Effect of Vitamin A deficiency (yes versus no)				
	Coef.	SE	OR	95 CI of OR	P-value
Marginal Logistic Model (exchangeable correlation)	0.276	0.223	1.32	0.85 to 2.04	0.217
Logistic Model with Random Intercept	0.658	0.426	1.93	0.84 to 4.45	0.123
Transition Model	0.156	0.161	1.17	0.85 to 1.60	0.335