Estimation of sensitivity and specificity of cervical cancer screening test

Raifu, Amidu[1]; Walter, Stephen[2]; Franco, Eduardo[3]; Agnihotram, Ramana-Kumar[3]

[1]Brock University; [2]McMaster University; [3]McGill University

GHSUG - Golden Horseshoe SAS Users Group

October 25, 2012

Outline of the Presentation

- Background of the Study
- Description of the data set
- What is Sensitivity and Specificity?
- Empirical and logistic regression estimation
- Results
- Conclusion

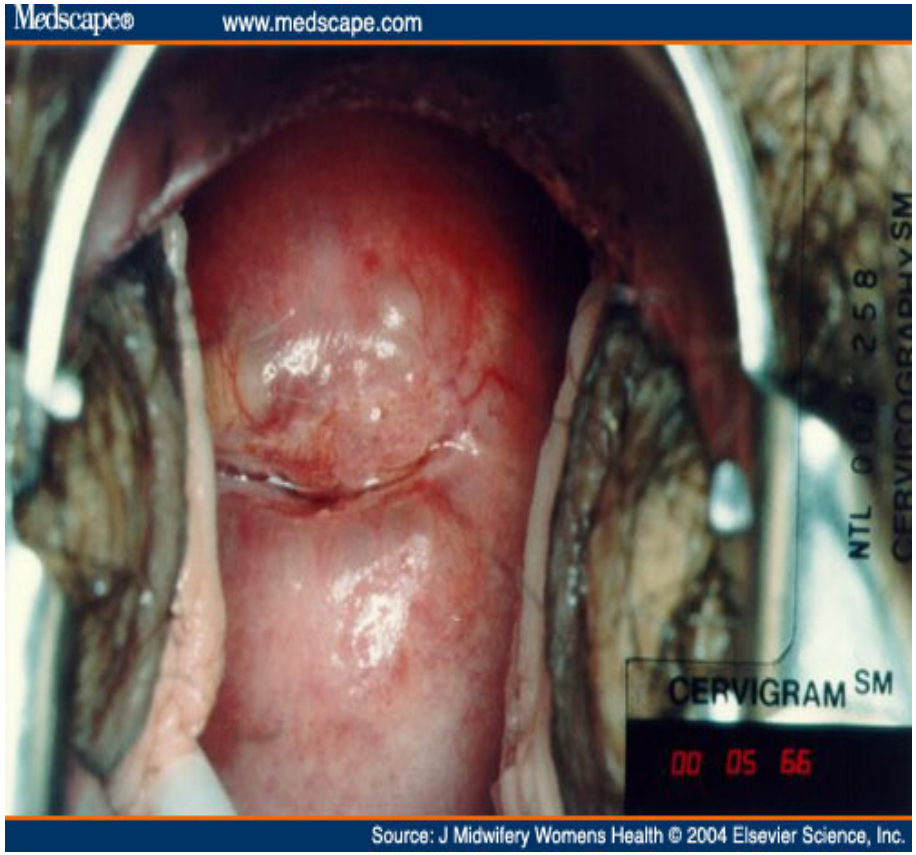
Background of the study

- The cervical cancer screening study
- Women aged 30 years old and above
- Visual Inspection with Acetic Acid (VIA) and with Lugol Iodine (VILI)
- Independent measurement by trained nurse and gynecologist

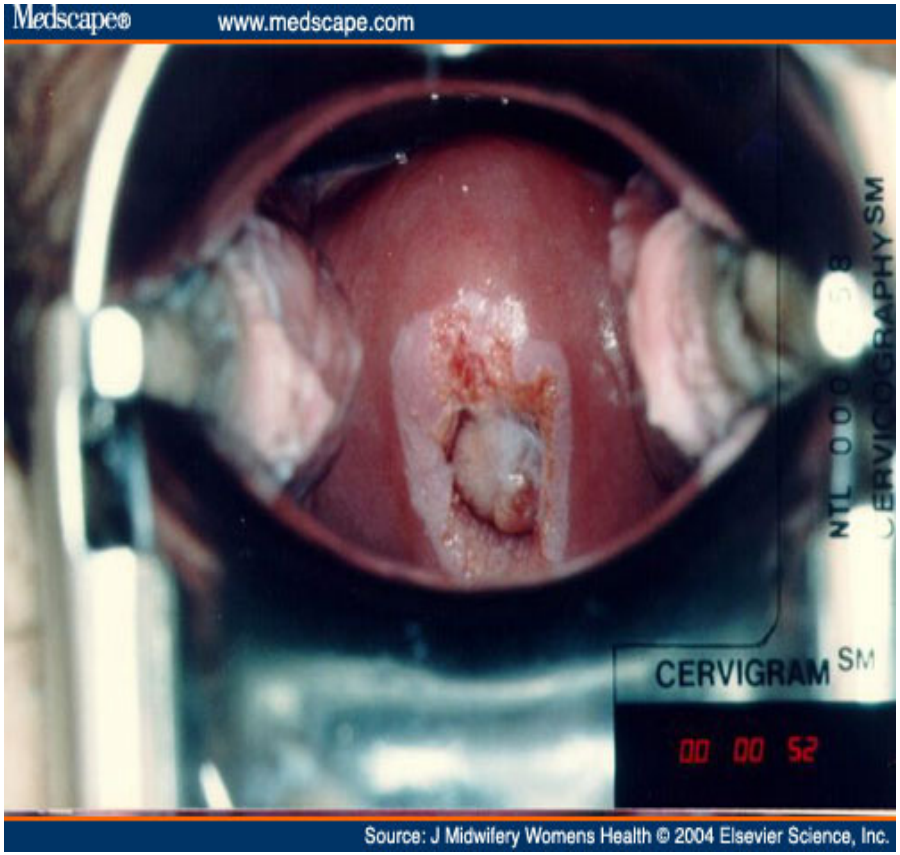
Map of Africa and DR Congo

VIA and VILI test application

VIA and VILI tests outcomes



Negative Outcome



Positive Outcome

Description of the data set

- 1,528 observations with 22 variables
- 1 Gold standard variable: *Histcolpo*
- Outcome variables:
  - VIA_nurse (positive vs. negative)
  - VIA_md (positive vs. negative)
- Covariates: age, parity, HC2, smoking, marital, etc

What is Sensitivity and Specificity?

- Sensitivity – Probability of a screening test being positive given that the disease is present

- Specificity – Probability of a screening test being negative given that the disease is not present

Empirical estimation

- Given a screening Test A and gold standard Test B

Screening Test A	Gold Standard Test B	
	+ve	-ve
+ve	a	b
-ve	c	d
Total	n_1	n_2

Sensitivity = P(Test A = +ve | Test B = +ve) = a/n_1 = TP/(TP + FN)

Specificity = P(Test A = -ve | Test B = -ve) = d/n_2= TN/(TN + FP)

Clopper-Pearson confidence interval (Exact binomial confidence intervals)

Logistic regression model

- Extended Coughlin et al's model (logistic regression)

$$\pi(\mathbf{x}) = P(\mathbf{Y} = 1 \mid \mathbf{X} = \mathbf{x}) = \frac{\exp\left(\beta_0 + \sum_{p=0}^{P} \beta_p X_p\right)}{1 + \exp\left(\beta_0 + \sum_{p=0}^{P} \beta_p X_p\right)}$$

$$Logit\left[\pi(\mathbf{x})\right] = \log\left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right] = \beta_0 + \sum_{p=0}^{P} \beta_p X_p$$

$$Logit\left[\pi(\mathbf{x})\right] = \beta_0 + \beta_1 * histcolpo + \beta_2 * smoking + \beta_3 * marital + \beta_4 * age + \beta_5 * parity +$$

$$\beta_6 * hc2 + \beta_7 * age * histcolpo + \beta_8 * parity * histcolpo + \hat{\beta}_9 * age^2 +$$

$$\beta_{10} * age^2 * histcolpo + \beta_{11} * parity^2 + \beta_{12} * parity^2 * histcolpo$$

Logistic regression estimation

$$Sensitivity = \frac{1}{1+\exp\left[-\left(\hat{\beta}_0 + \sum_{p=0}^{P}\hat{\beta}_p X_p\right)\right]}$$

where histcolpo = 1

$$Sensitivity = \frac{1}{1+\exp\left[-\left(\begin{array}{l}\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2*smoking+\hat{\beta}_3*marital+\left(\hat{\beta}_4+\hat{\beta}_7\right)*age+\\ \left(\hat{\beta}_5+\hat{\beta}_8\right)*parity+\hat{\beta}_6*hc2+\left(\hat{\beta}_9+\hat{\beta}_{10}\right)*age^2+\\ \left(\hat{\beta}_{11}+\hat{\beta}_{12}\right)*parity^2\end{array}\right)\right]}$$

$$Specificity = 1-\left\{\frac{1}{1+\exp\left[-\left(\hat{\beta}_0 + \sum_{p=0}^{P}\hat{\beta}_p X_p\right)\right]}\right\}$$

where histcolpo = 0

$$Specificity = 1-\left\{\frac{1}{1+\exp\left[-\left(\begin{array}{l}\hat{\beta}_0+\hat{\beta}_2*smoking+\hat{\beta}_3*marital+\hat{\beta}_4*age+\hat{\beta}_5*parity+\\ \hat{\beta}_6*hc2+\hat{\beta}_9*age^2+\ \hat{\beta}_{11}*parity^2\end{array}\right)\right]}\right\}$$

Results (Empirical method)

Empirically estimated sensitivity and specificity of VIA test measured by nurse

Age group(y)	Sensitivity	95% Confidence Interval	Specificity	95% Confidence Interval
30-34	0.67	(0.38, 0.88)	0.51	(0.44, 0.57)
35-39	0.69	(0.41, 0.89)	0.67	(0.60, 0.73)
40-44	1.00	(0.78, 1.00)	0.63	(0.56, 0.69)
45-49	0.86	(0.42, 0.99)	0.73	(0.67, 0.78)
50-54	0.71	(0.29, 0.96)	0.70	(0.63, 0.77)
55-59	0.71	(0.29, 0.96)	0.77	(0.68, 0.85)
60-64	0.75	(0.19, 0.99)	0.82	(0.70, 0.90)
65-69	1.00	(0.03, 1.00)	0.56	(0.38, 0.74)
70+	0.60	(0.15, 0.95)	0.64	(0.48, 0.78)

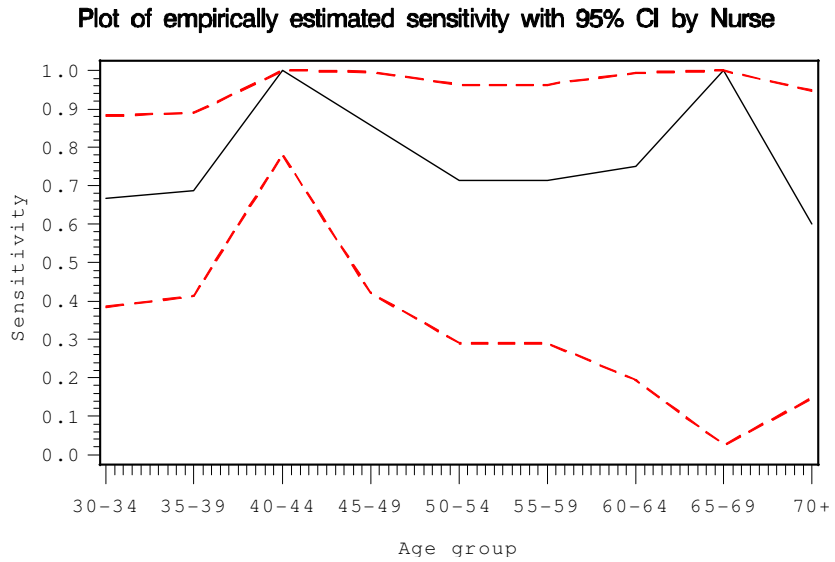
Results (Empirical method)

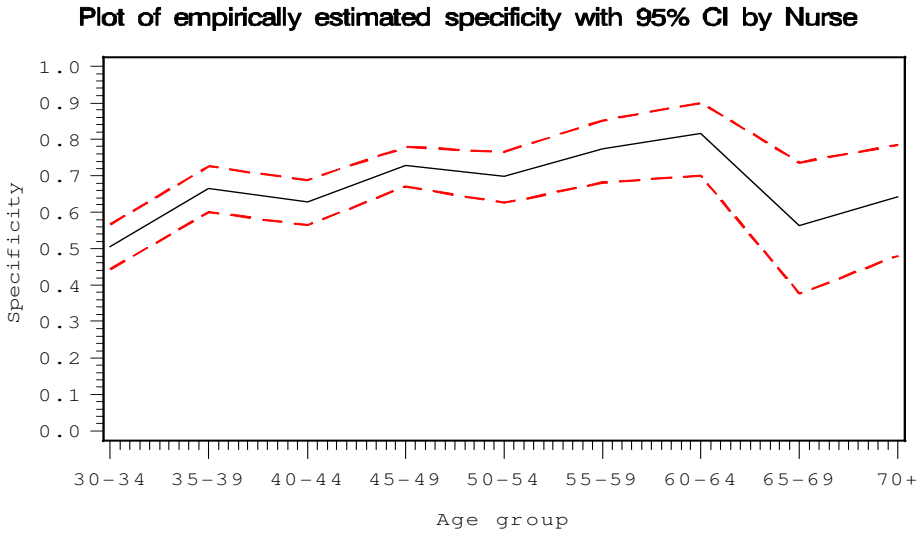
Empirically estimated sensitivity and specificity of VIA test measured by physician

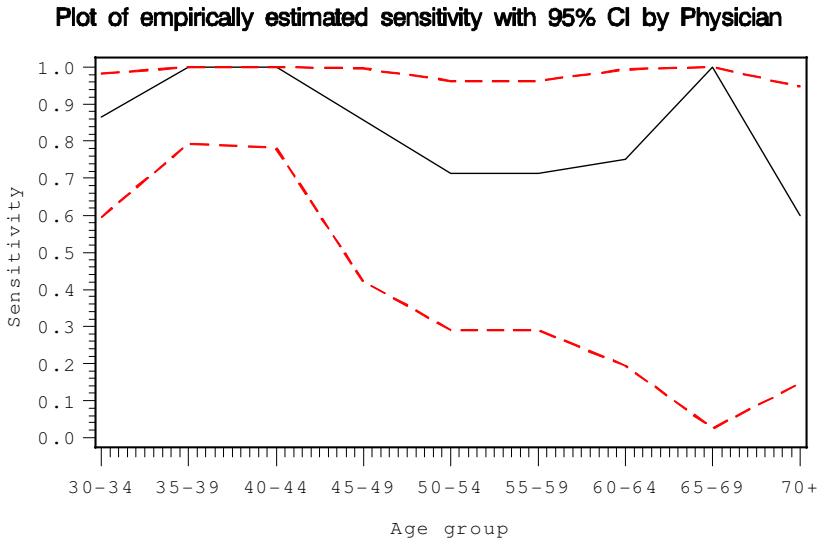
Age group(y)	Sensitivity	95% Confidence Interval	Specificity	95% Confidence Interval
30-34	0.87	(0.60, 0.98)	0.59	(0.53, 0.65)
35-39	1.00	(0.79, 1.00)	0.73	(0.67, 0.79)
40-44	1.00	(0.78, 1.00)	0.72	(0.66, 0.78)
45-49	0.86	(0.42, 0.99)	0.77	(0.71, 0.82)
50-54	0.71	(0.29, 0.96)	0.82	(0.76, 0.88)
55-59	0.71	(0.29, 0.96)	0.79	(0.70, 0.87)
60-64	0.75	(0.19, 0.99)	0.82	(0.70, 0.90)
65-69	1.00	(0.03, 1.00)	0.59	(0.41, 0.76)
70+	0.60	(0.15, 0.95)	0.76	(0.61, 0.88)

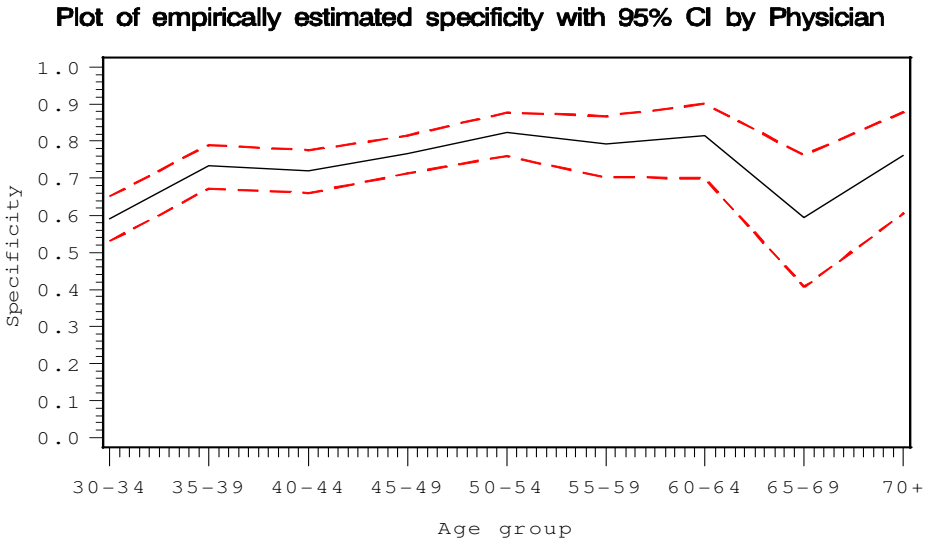
Results (Empirical method)

Plots of empirically estimated sensitivity and specificity of VIA test









Results (Logistic regression)

Outcome variable	Parameter	Estimate	P-value	95% Confidence Interval
via_nurse	Intercept	-0.7829	< 0.0001	(-0.9653, -0.6004)
	age	-0.0416	< 0.0001	(-0.0557, -0.0275)
	parity	0.0348	0.0933	(-0.0058, 0.0755])
	histcolpo	1.8956	< 0.0001	(1.1831, 2.6082)
	age^2	0.0021	< 0.0001	(0.0011, 0.0030)
	$histcolpoage^2$	-0.0018	0.2293	(-0.0047, 0.0011)
	$parity^2$	-0.0094	0.0424	(-0.0184, -0.0003)
via_md	Intercept	-1.2808	< 0.0001	(-1.4598, -1.1019)
	age	-0.0436	< 0.0001	(-0.0587, -0.0285)
	parity	0.0489	0.0248	(0.0062, 0.0915)
	histcolpo	2.8610	< 0.0001	(2.0940, 3.6282)
	age^2	0.0021	< 0.0001	(0.0011, 0.0031)
	HC2	0.4257	0.0262	(0.0504, 0.8011)

Results (Logistic regression)

Model-based estimated sensitivity and specificity of VIA test measured by nurse

Age group(y)	Sensitivity	95% Confidence Interval	Specificity	95% Confidence Interval
30-34	0.81	(0.68, 0.90)	0.52	(0.48, 0.57)
35-39	0.80	(0.66, 0.89)	0.61	(0.57, 0.65)
40-44	0.76	(0.61, 0.87)	0.67	(0.63, 0.71)
45-49	0.73	(0.57, 0.85)	0.72	(0.68, 0.76)
50-54	0.70	(0.52, 0.83)	0.73	(0.69, 0.77)
55-59	0.65	(0.47, 0.79)	0.74	(0.70, 0.77)
60-64	0.61	(0.42, 0.76)	0.72	(0.68, 0.76)
65-69	0.59	(0.40, 0.75)	0.67	(0.63, 0.72)
70+	0.53	(0.35, 0.70)	0.62	(0.57, 0.66)

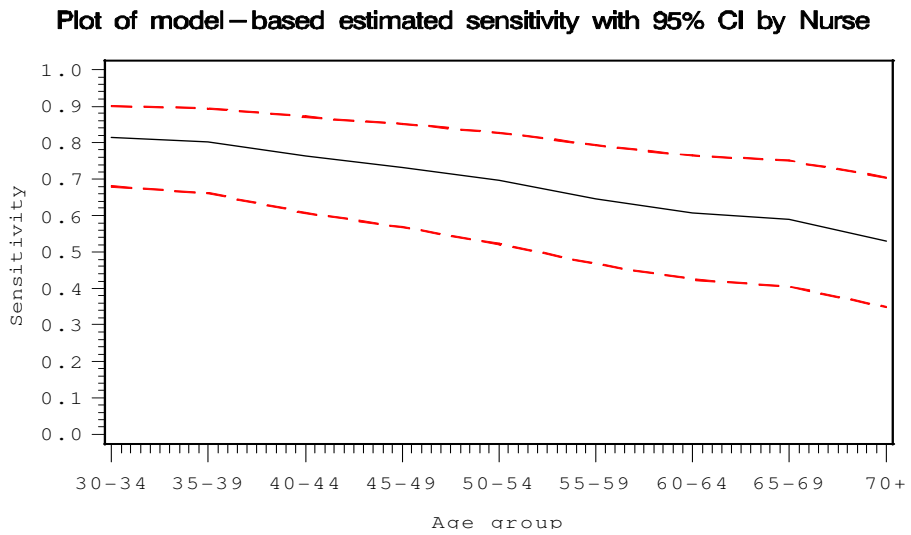
Results (Logistic regression)

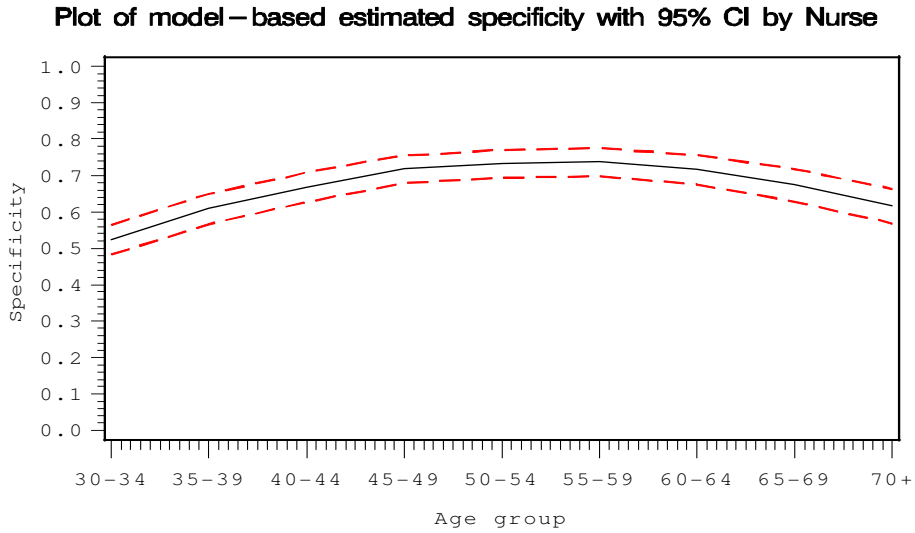
Model-based estimated sensitivity and specificity of VIA test measured by physician

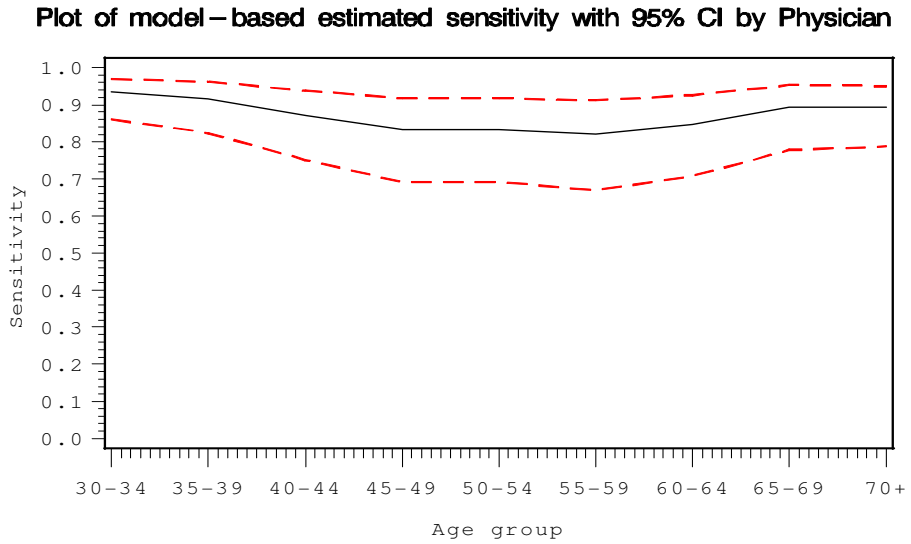
Age group(y)	Sensitivity	95% Confidence Interval	Specificity	95% Confidence Interval
30-34	0.93	(0.86, 0.97)	0.60	(0.56, 0.65)
35-39	0.92	(0.82, 0.96)	0.69	(0.65, 0.73)
40-44	0.87	(0.75, 0.94)	0.74	(0.70, 0.78)
45-49	0.83	(0.69, 0.92)	0.78	(0.74, 0.81)
50-54	0.83	(0.69, 0.92)	0.79	(0.76, 0.83)
55-59	0.82	(0.67, 0.91)	0.80	(0.77, 0.83)
60-64	0.85	(0.71, 0.93)	0.79	(0.75, 0.82)
65-69	0.89	(0.78, 0.95)	0.75	(0.71, 0.79)
70+	0.89	(0.79, 0.95)	0.72	(0.67, 0.75)

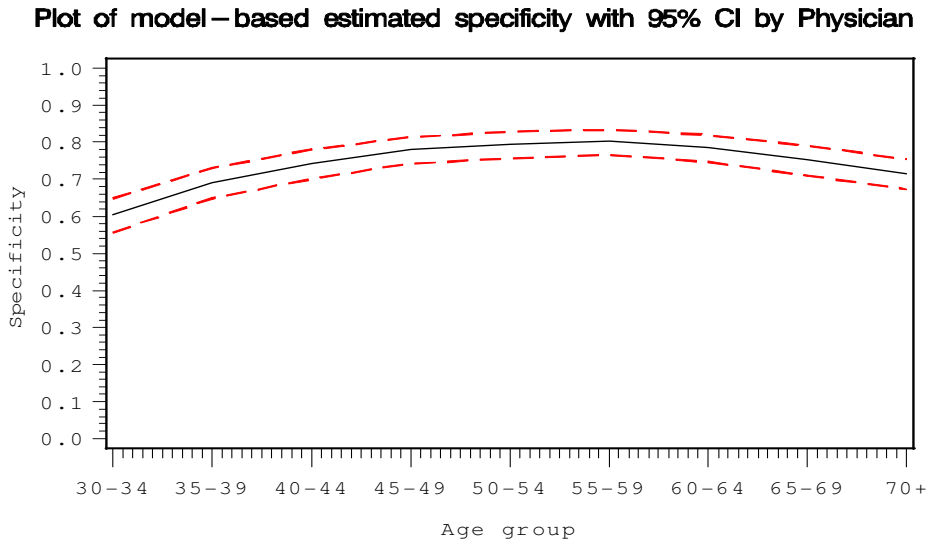
Results (cont'd)

Plots of model-based estimated sensitivity and specificity of VIA test across age group









Conclusions

- Age, parity, and HPV DNA influenced the sensitivity and specificity of VIA test

- These results will guide the investigators when conducting similar study in the future

SAS codes

Proc freq procedure

```
title ‘Empirical estimation of sensitivity of VIA by nurse’;
proc freq data = viatest  order = freq;
   by agegrp;
   where histcolpo= 1;
   tables vianurse;
   exact binomial;
   output out=sens_cl n binomial;
run;
```

Specifies the order for reporting variable values

Requests exact tests or confidence limits for the specified statistics

Request for Binomial proportion tests for one-way tables

Proc logistic procedure

```
proc logistic data = viatest;
   class vianurse hc2(ref='0')/param=ref ;
   model vianurse (event='1') =  agec paritc histcolpo agec2
   histagec2 paritc2/covb lackfit;
   output out=diag1 prob=p1 lower=LCL upper=UCL;
run;
```

Specify goodness of fit test

Specifies the variance-covariance matrix for the parameters in the model

SAS codes

```
/*Model Estimate for Sensitivity*/
data sen1; set diag1;
sens = 1/(1 + exp(-(-0.7829 - 0.0416*agec + 0.0348*paritc +
   1.8956*histcolpo1 + 0.00208*agec2 - 0.00181*histagec2 -
   0.00935*paritc2)));

sensUCL = 1/(1 + exp(-(-0.7829 - 0.0416*agec + 0.0348*paritc +
   1.8956*histcolpo1 + 0.00208*agec2 - 0.00181*histagec2-0.00935*paritc2)
   - 1.96*(sqrt(0.008667+0.000052*agec+0.00043*paritc+0.132179*histcolpo1
   + 2.355E-7*agec2+2.256E-6*histagec2
      +0.000021*paritc2))));

sensLCL = 1/(1 + exp(-(-0.7829 - 0.0416*agec + 0.0348*paritc +
   1.8956*histcolpo1 + 0.00208*agec2 - 0.00181*histagec2-0.00935*paritc2)
   +1.96*(sqrt(0.008667+0.000052*agec+0.00043*paritc+0.132179*histcolpo1+
   2.355E-7*agec2+2.256E-6*histagec2
      +0.000021*paritc2))));
where histcolpo = 1;
run;
```

Proc means procedure

```
proc means data = sen1 mean clm  std;
   var sens sensLCL sensUCL;
   class agegrp;
   output out=sens1 mean=mean(sens sensLCL sensUCL) = ;
run;
```

Requests for statistics to be included in the output data set

SAS codes

```
goptions reset=global htitle=1.3;
axis1 major=(h=1) minor=(h=0.5)
width=3 value=(h=1.2)
label=(color=black h=1.2 "Age group");
axis2 order=(0 to 1 by 0.1) major=(h=1)
minor=(h=0.5) width=3 value=(h=1.2)
label=(angle=90 color=black h=1.2);
proc gplot data = vianursediag;
plot sens*agegrp
   (senLCL senUCL)*agegrp/ overlay haxis=axis1 vaxis=axis2;
   symbol1 c=black i=join w=2 l=1;
   symbol2 c=red i=join w=2  l=3;
   symbol3 c=red i=join w=2 l=3;
   label sens="Sensitivity";
   title "Plot of empirically estimated sensitivity with 95% CI by Nurse";
   run;

reset=global   Reset all goptions to default value
htitle = # Specify the height of the title
major=(h=#) Specify the height of the major marks on the axes
minor=(h=#) Specify the height of the minor marks on the axes
width=#  Specify the thickness of the borderline on the graph
angle=# Turns the entire line of the text the number of degree
c = color of the symbol; i= join the graph points
w = specify the thickness of the graph lines
l = specify the type of line
```

References

- Coughlin SS, Trock B, Criqui MH, et al. The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test. *Journal of Clinical Epidemiology* 1992; **45**(1):1-7

- Walter SD. Estimation of Test Sensitivity and Specificity when Disease Confirmation is Limited to Positive Results. *Epidemiology* 1999; **10**(1): 67-72

- Sangwa-Lugoma G, Mahmud S, Nasr SH, et al. Visual inspection as a cervical cancer screening method in a primary health care setting in Africa. *International Journal of Cancer* 2006; **119**:1389-1395

Thank you !