Comparing effects across nested logistic regression models

CAPS Methods Core
Quantitative Working Group Seminar

September 23, 2011

Steve Gregorich

TABLE 3. Crude and AORs for Maternal-to-Child Transmission Comparing ZDV-Sparing With ZDV-Containing HAART in Pregnancies

	Univariable Model (Numbers Vary Due to Missing Data)			Multivariable Model (n = 6111)*		
	n (%)	OR (95% CI)	P	n (%)	AOR (95% CI)	P
ART group			0.64			0.18
ZDV-containing HAART	5227 (0.9)	1		5214 (0.9)	1	
ZDV-sparing HAART	903 (0.8)	0.83 (0.37 to 1.83)		897 (0.8)	1.81 (0.77 to 4.26)	
Study			0.68			0.72
NSHPC	5261 (0.9)	1		5247 (0.9)	1	
ECS	869 (1.0)	1.16 (0.57 to 2.38)		864 (0.9)	1.15 (0.53 to 2.47)	
Mode of delivery						
Elective CS	3515 (0.8)	1	—	3515 (0.8)	1	—
Emergency CS	1095 (1.7)	2.28 (1.26 to 4.12)	<0.01	1095 (1.7)	2.07 (1.13 to 3.76)	0.02
Vaginal	1051 (0.9)	0.78 (0.37 to 1.66)	0.52	1051 (0.9)	0.80 (0.38 to 1.72)	0.58
Duration of HAART (wks)						
≥24	2087 (0.2)	1	—	2078 (0.2)	1	—
12–23	2416 (0.7)	2.95 (1.09 to 8.01)	0.03	2410 (0.7)	3.44 (1.20 to 9.86)	0.02
8–11	1020 (1.0)	4.12 (1.41 to 12.09)	0.01	1017 (1.0)	5.10 (1.65 to 15.77)	0.01
2–7	607 (4.0)	17.14 (6.51 to 45.12)	<0.001	606 (4.0)	20.09 (7.22 to 55.93)	<0.001

*Adjusted for study, mode of delivery, and duration of HAART.
ART, antiretroviral therapy; CS, cesarean section; OR, odds ratio.

Comparing parameter estimates across two nested linear models

Covariate-adjusted (*Full*) model

$$\dot{y}_i = a_{\mathrm{F}} + x_i\dot{b}_{x.\mathrm{F}} + c_i\dot{b}_{c.\mathrm{F}} + \dot{e}_{i.\mathrm{F}}$$

Unadjusted (*Restricted*) model

$$\dot{y}_i = a_{\mathrm{R}} + x_i\dot{b}_{x.\mathrm{R}} \qquad\qquad + \dot{e}_{i.\mathrm{R}}$$

What is the effect of adjustment for c?

. Compare $\dot{b}_{x.\mathrm{F}}$ to $\dot{b}_{x.\mathrm{R}}$, either formally or just 'eyeball' the difference

Comparing parameter estimates across two nested logistic models

Covariate-adjusted (*Full*) model

$$\text{logit}(y_i = 1 | x_i, c_i) = a_{\text{F}} + x_i b_{x.\text{F}} + c_{\text{i}} b_{c.\text{F}}$$

Unadjusted (*Restricted*) model

$$\text{logit}(y_i = 1 | x_i) = a_{\text{R}} + x_i b_{\text{x.R}}$$

. Here, comparing $b_{x.\text{F}}$ to $b_{x.\text{R}}$ is more complex

. To understand why, we'll look at the binary outcome threshold model

Binary outcome regression represented as a threshold model

. y^* is an unobserved (latent) continuous outcome variable representing the propensity of outcome occurrence

$$y_i^* = \underset{\cdot}{a} + x_i \underset{\cdot}{b} + \underset{\cdot}{e}_i,$$

where $\underset{\cdot}{e}_i$ ~ Logistic(0,π^2/3) for logistic or N(0,1) for probit

. Usually, the relationship between continuous y^* and binary y is defined as

if $y_i^* > 0$ then $y_i = 1$;

else $y_i = 0$

Given $\underset{\cdot}{e}_i$ ~ Logistic(0,π^2/3), model parameters for <u>correctly specified</u> models are equivalent across

linear model of y^*, and

logistic model of y

Three identifying assumptions of logistic regression model

. conditional mean of $e_i = 0$

. $\text{Var}(e_i|x) = \pi^2/3$

. threshold value for y^* is 0 (usually):

if $y^* > 0$ then y = 1;

else y = 0

Comparing linear and logistic regression

Basics of modeled variation

	outcome variance	residual variance
linear regression (y)	σ_y^2 is observed	$\sigma_{\dot{e}}^2$ is model-dependent
logistic regression	$\sigma_{y^*}^2$ is model-dependent	$\sigma_{\underset{\cdot}{e}}^2$ is fixed

Effects of added X variables on modeled variation

	outcome variance	residual variance
linear regression (y)	σ_y^2 unchanged	$\sigma_{\dot{e}}^2$ decreased
logistic regression (y*)	$\sigma_{y^*}^2$ increased	$\sigma_{\underset{\cdot}{e}}^2$ unchanged

. Adding explanatory vars. to a logistic model, increases implied variance of y*

. Essentially, y* is rescaled.

. When y* is rescaled, model parameters are also rescaled. Same for models of y

Comparing parameters across nested logistic regression models

$\text{logit}(y_i = 1 | x_i, c_i) \quad = a_{\text{F}} + x_i b_{x.\text{F}} + c_i b_{c.\text{F}}$ (*Full* model)

$\text{logit}(y_i = 1 | x_i) \quad = a_{\text{R}} + x_i b_{x.\text{R}}$ (*Restricted* model)

. $b_{1.\text{F}}$ and $b_{1.\text{R}}$ may differ because of

. confounding (expectation: $b_{x.\text{F}} < b_{x.\text{R}}$)

. negative confounding (expectation: $b_{x.\text{F}} > b_{x.\text{R}}$)

. rescaling (expectation: $b_{x.\text{F}} > b_{x.\text{R}}$)

. a combination (expectation: ??)

Parameter rescaling is almost universally unknown/ignored except in specific contexts

. testing mediation

. generalized linear mixed models

A simulated example of faux negative confounding

Simulated data

. A single sample with N=500,000

. x and c are bivariate normal with the following sample statistics (exactly)

. $\bar{x} = \bar{c} = 0$

. $\sigma_x^2 = 1, \quad \sigma_c^2 = 4$

. $r_{xc} = 0$

. Next, I used x and c values to generate a continuous y^* variate as

$y_i^* = x_i + c_i + e_i,$ (i.e., both regression parameters equaled unity)

where the e_i ~ Logistic(0,π^2/3)

. Finally, I created a binary version of y^* as

$y = 1$ if $y^* > 0$;

$y = 0$ otherwise

A simulated example of faux negative confounding

Results of linear models regressing y^* onto x and c

	Full model Adjusted b	*Restricted* model Unadjusted b
x modeled, c excluded	1.00	1.00
c modeled, x excluded	1.00	1.00

Results of logistic models regressing y onto x and c

	Full model Adjusted b	*Restricted* model Unadjusted b
x modeled, c excluded	1.00	0.61
c modeled, x excluded	1.00	0.85

A simulated example of faux negative confounding

Explanation for results on previous slide

In this simplified example, x and c are orthogonal,
so the implied variance of y^* equals

Full model

$$\sigma^2_{y^*.\mathrm{F}} = \sigma^2_x b^2_{x.\mathrm{F}} + \sigma^2_c b^2_{\mathrm{c.F}} + \pi^2/3 = 8.29$$

Restricted model including x

$$\sigma^2_{y^*.\mathrm{R}} = \sigma^2_x b^2_{x.\mathrm{R}} + \pi^2/3 = 3.66$$

Scaling of the outcome and parameter estimates is not equivalent across models

One attempted solution in the literature

. In the context of testing mediation, Winship and Mare (1984) and MacKinnon & Dwyer (1993) suggested a rescaling of model parameters based upon the $\sigma^2_{y^*.\mathrm{F}}$ and $\sigma^2_{y^*.\mathrm{R}}$ to allow comparison of, e.g., $b_{x.\mathrm{R}}$ and $b_{x.\mathrm{F}}$

. This is known as *y*-standardization.
However, it does not work very well

. For the previous example, the rescaled value of $b_{x.\mathrm{R}}$ equals

$$\text{rescaled } b_{x.\mathrm{R}} = 0.61 \times \sqrt{\frac{\sigma^2_{y^*.\mathrm{F}}}{\sigma^2_{y^*.\mathrm{R}}}} = 0.61 \times \sqrt{\frac{8.29}{3.66}} = 0.61 \times 1.51 = 0.92, \text{ not } 1.00$$

. There have been other proposed solutions that I have not studied
(reportedly they don't work well, either)

Karlson, Holm, & Breen (KHB) (in press)

. KHB argue that the scaling is a factor of the error standard deviation, σ_e, not the standard deviation of y^*

. Of course y^* and σ_e are unobserved, in practice, but given our simulated data, we can take a look

. For the *Full* model, $\sigma^2_{e.\mathrm{F}} = \pi^2/3 = 3.29$

. For the *Restricted* model, $\sigma^2_{e.\mathrm{R}} = \sigma^2_c + \pi^2/3 = 7.29$

. Therefore, the KHB-suggested rescaled value equals

$$\boxed{\text{rescaled } b_{x.\mathrm{R}} = 0.61 \times \sqrt{\frac{\sigma^2_{e.\mathrm{R}}}{\sigma^2_{e.\mathrm{F}}}} = 0.61 \times \sqrt{\frac{7.29}{\pi^3/3}} = 0.61 \times 1.49 = 0.91, \text{ not } 1.00}$$

Comparing parameter rescaling methods

From the earlier simulated example

	$\hat{b}$			
	Full model	*Restricted* model	*Restricted* σ_{y^*}-rescaled	*Restricted* σ_e-rescaled
x	1.00	0.61	0.92	0.91

Regardless of these results, KHB suggest a method to rescale parameter estimates from binary outcome models that appears to work.

KHB method

. Here, C_i refers to the vector of covariates in the *Full* model

. Replace all covariates, C_i, in the *Full* logisitc regression model with
residuals from regression of C_i on *x*, R_i. Name this the *KHB* model

. The *KHB* model provides an estimate of the unadjusted effect of *x* on *y*
that is on the same scale as parameters from the *Full* model

. Clever

The R_i are uncorrelated with *x*

The *KHB* model obtains an unadjusted estimate of the *x* effect.
(the *KHB* model obtains Type 1 estimates of the *x* effect).

model-dependent σ_{y^*} and σ_e are equivalent across the *KHB* and *Full* models

The *KHB* model obtains unadjusted parameter estimates for *x*
that are on the scale of the *Full* model.

. Method easily extends to accommodate any number of *x* and *c* variables

KHB method

What about binary covariates?

KHB suggest using the linear probability model (LPM)
to generate residuals of the C_i

Then fit the *KHB* model in the usual way

KHB method

LPM

. Fit a linear regression model of the binary outcome

. Conditional expectation of *y* given *x*, $E(y_i|x_i) = \Pr(y_i\!=\!1|x) = a + x_i b_x + c_i b_c$

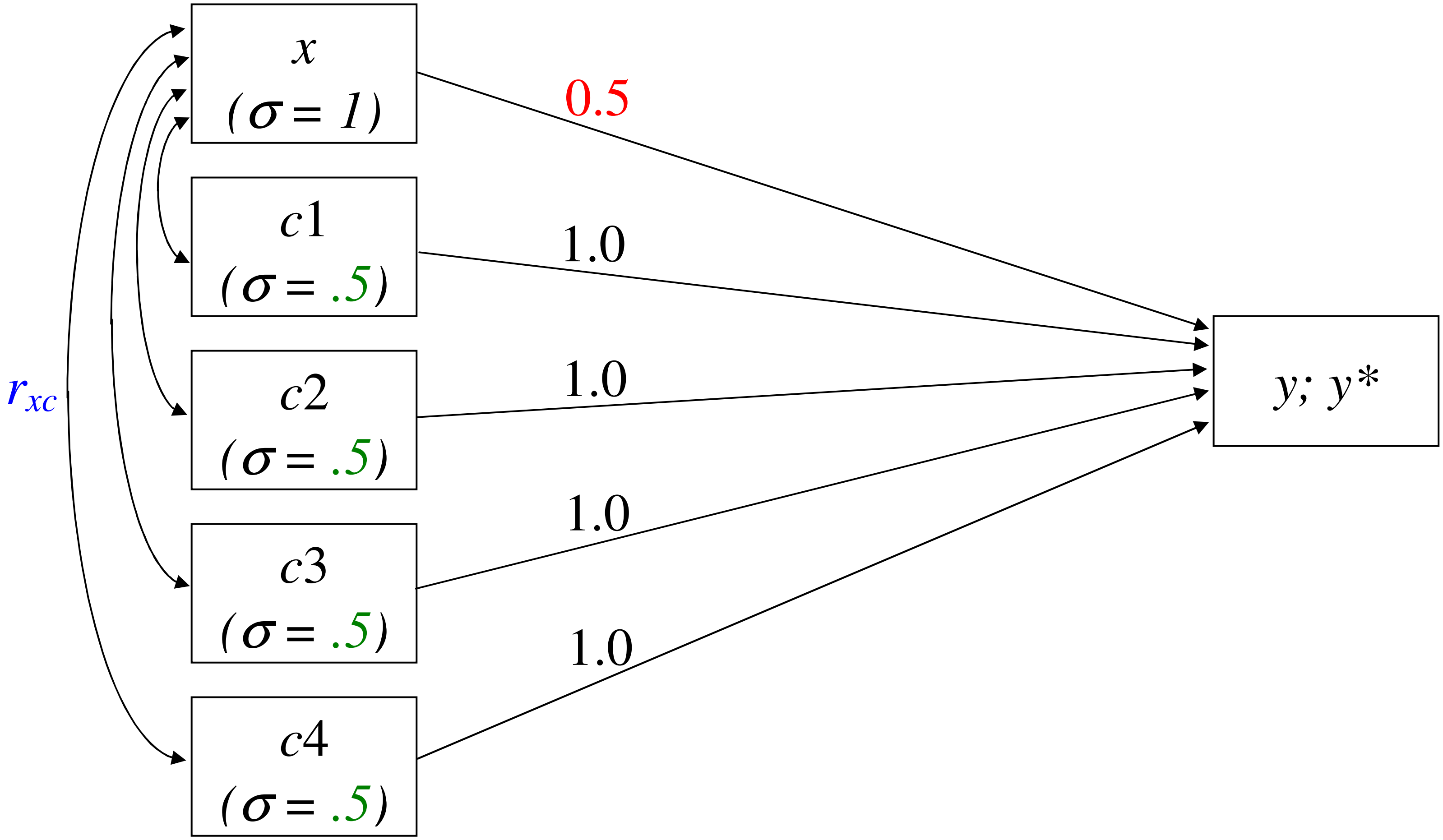
. Binary *y* does not affect interpretation the parameters, compared to continuous *y*.
For a unit change in *x*, the expected change in the probability that *y*=1 is b_x, holding any control variables constant.

. Because the model is linear, a unit change in *x* always results in the same change in probability—the model is linear in the probability.

. In general practice, there are problems with the linear probability model:
. heteroskedasticity (the variance of *y*|*x* depends on *x*)
. residuals cannot be normally distributed
. predicted probabilities outside [0,1]
. functional form

Even so, the LPM could meet the needs of the *KHB* model
I.e., to estimate the unadjusted effect of *x* on the scale of the *Full* model

Simulation study: Population model



Unadjusted effects of x ($b_{.R}$) as a function of r_{xc}:

. $r_{xc} = 0.250$; $b_{.R} = 0.5 + 0.250 \times 0.5 \times 4 = 1.00$

. $r_{xc} = 0$; $b_{.R} = 0.5 + 0 \times 0.5 \times 4 = 0.50$

. $r_{xc} = -0.125$; $b_{.R} = 0.5 + -0.125 \times 0.5 \times 4 = 0.25$

Simulation Details

. N=125, 250, 500, <u>1000</u>

. R=1000

. $x \sim N(0, 1)$

. $c1$ - $c4$ ~$N(0, 0.25)$; or $B(0.50)$. *2 conditions: norm./bin. c; variance = 0.25*

. $b_x = b_{\cdot x} = 0.5$; $b_c = b_{\cdot c} = 1.0$

. $r_{xc} = 0.25$; 0; –0.125. *3 conditions: pos., no, and neg. confounding*

. $r_{cc} = 0$

$y_i^* = x_i 0.5 + c1_i + c2_i + c3_i + c4_i + e_{\cdot i}$, where $e_{\cdot i} \sim \text{Logistic}(0, \pi^2/3)$

if $y_i^* > 0$ then $y_i = 1$; else $y_i = 0$

. $y^* \sim N(0, †)$

. $y \sim B(0.50)$

† dependent on r_{xc}: ranges from approximately 4.0 to 5.0

Simulation results: N=1000. R=1000 replicate samples

Continuous x and c:

	linear reg: y^*		logistic reg: y		KHB: y			
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
r_{xc}	$\widehat{b}_{\cdot R}$	$\frac{\hat{\sigma}_e}{\sqrt{\pi^2/3}}$	$\widehat{b}_R$	(b)×(c)	$\widehat{b}_{\cdot\cdot R}$	$\overline{\widehat{se}}$	$\sigma_{\widehat{b}_R}$	covg.
+0.25	**1.00**	**1.13**	**0.84**	**0.96**	**1.00**	**0.09**	**0.09**	**0.933**
0.0	**0.50**	**1.14**	**0.41**	**0.47**	**0.50**	**0.07**	**0.08**	**0.929**
–0.125	**0.25**	**1.14**	**0.20**	**0.23**	**0.25**	**0.07**	**0.08**	**0.932**

Continuous x and binary c:

	linear reg: y^*		logistic reg: y		KHB: y			
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
r_{xc}	$\widehat{b}_{\cdot R}$	$\frac{\hat{\sigma}_e}{\sqrt{\pi^2/3}}$	$\widehat{b}_R$	(b)×(c)	$\widehat{b}_{\cdot\cdot R}$	$\overline{\widehat{se}}$	$\sigma_{\widehat{b}_R}$	covg.
+0.25	**1.00**	**1.11**	**0.91**	**1.00**	**1.02**	**0.11**	**0.11**	**0.953**
0.0	**0.50**	**1.14**	**0.43**	**0.49**	**0.51**	**0.09**	**0.09**	**0.925**
–0.125	**0.25**	**1.13**	**0.21**	**0.24**	**0.25**	**0.08**	**0.09**	**0.928**

Some implications about naïve point estimates of b_R

If you naïvely compare b_F to b_R, you might draw incorrect conclusions

Results for continuous x and c

r_{xc}	b_F	$\widehat{b_R}$ naïve	$\widehat{b_R}$ KHB	$\Delta_{naïve}$	Δ_{true}
+0.250	0.50	0.91	1.00	+0.41†	+0.50
0	0.50	0.43	0.50	–0.07‡	0
–0.125	0.50	0.21	0.25	–0.29*	–0.25

† under-estimating the degree of positive confounding

‡ suggesting negative confounding when none exists

* over-estimating the degree of negative confounding

Simulations were simplistic

. models with multiple covariates may include those that are positively, negatively, and un-confounded with x

More

Tests of differences between adjusted and rescaled unadjusted effects

Normally I don't care about this (except in the context of testing mediation)

KHB present a test based upon Sobel.

Can accommodate multiple x and multiple c variables

Known problems with Sobel, Aroian, etc

Conclusions

. KHB model is simple to implement

. Quality of KHB model point estimates

Seems to do a good job of obtaining rescaled unadjusted point estimates

Use of LPM for binary covariates seemed to work well

I considered other scenarios,

Varied the distribution of binary *c* and *y*

Lognormal distribution of X

KHB (2011) report upon a fairly extensive simulation study

. Quality of KHB model standard errors/coverage

Coverage of rescaled unadjusted *x* effects was just OK

in my limited simulation.

If one wants to emphasize any tests of rescaled unadjusted effects,

the bootstrap should be considered

Resources

KHB papers (contact Kristian Karlson: kbk@sfi.dk)

1. Kristian Bernt Karlson, Anders Holm, and Richard Breen. (March, 09, 2011). *Comparing Regression Coefficients Between Models using Logit and Probit: A New Method*. Draft manuscript.

2. Kohler, U., Karlson, K.B., Holm, A. (in press). Comparing coefficients of nested nonlinear probability models. The Stata Journal.

3. Breen, R., Karlson, K.B., Holm, A. (April 11, 2011). Total, Direct, and Indirect Effects in Logit Models. Abstract available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1730065

4. Karlson, K.B. and Holm, A. (2011). Decomposing primary and secondary effects: A new decomposition method. *Research in Social Stratification and Mobility*, *29*, 221-237. http://www.sciencedirect.com/science/article/pii/S0276562410000697

KHB Stata ado

http://ideas.repec.org/c/boc/bocode/s457215.html