Comparing effects across nested logistic regression models

CADC Scholars Meeting

March 12, 2013

Steve Gregorich
Example from the literature of nested model comparisons

Care is required when comparing ORs across nested models
Outline

. Comparing Parameter Estimates Across Nested Linear Models

. Example Data

. Example Application: Linear Modeling Framework

. Example Application: Logistic Modeling Framework

. Reconciliation

. Conclusions

. Resources
Comparing parameter estimates across nested linear models

Initial concepts regarding linear regression

\[ y_i = \text{intercept} + x_i b + e_i \]

The total variance of \( y \) is decomposed into
- the variance explained by \( x \), plus
- residual (unexplained) variance

\[ \text{VAR}(y_i) = \text{VAR}(x_i) \times b^2 + \text{VAR}(e_i) \]

When you add additional \( x \) variables to the model,
- the explained variation increases and
- the residual variation decreases

\text{ALWAYS: total variation = explained + residual variation}
Comparing parameter estimates across nested linear models

Two models are nested if the parameters of one are a subset of the other

Unadjusted model: \( y_i = \text{intercept} + x_i b_U + e_i \)

Adjusted model: \( y_i = \text{intercept} + x_i b_A + \text{cov}_i b + e_i \)

The Unadjusted model is nested within the Adjusted model

What effect does adjustment for cov have on the modeled effect of \( x \)?

Compare \( b_A \) to \( b_U \), either formally or just 'eyeball' the difference

Unadjusted vs. Adjusted models are just one type of nested model comparison
Example data (L Karliner)

*N*=8077 patient discharges from UCSF 14th Floor
January 2007 - January 2010

**Main explanatory variable**
Binary patient sex indicator (MalePt): 49% male

**Additional $x$ variable**

<table>
<thead>
<tr>
<th>scale</th>
<th>Mean</th>
<th>Median</th>
<th>Variance</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOS (days)</td>
<td>5.5</td>
<td>3.7</td>
<td>32.6</td>
<td>0.1</td>
<td>39.8</td>
</tr>
<tr>
<td>$ln$LOS</td>
<td>1.3</td>
<td>1.3</td>
<td>0.7</td>
<td>-2.8</td>
<td>3.7</td>
</tr>
</tbody>
</table>

. LOS = length of (hospital) stay

**Outcome: Total Costs**

<table>
<thead>
<tr>
<th>scale</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costs ($)</td>
<td>19,073</td>
<td>10,483</td>
<td>1,335</td>
<td>393,304</td>
</tr>
<tr>
<td>$ln$Costs</td>
<td>9.37</td>
<td>9.26</td>
<td>7.20</td>
<td>12.88</td>
</tr>
<tr>
<td>binaryCosts</td>
<td>0.50</td>
<td>--</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Example Application: *Linear Model*

Modeling the effect of patient sex on *ln* Costs

<table>
<thead>
<tr>
<th></th>
<th>Unadjusted Model</th>
<th></th>
<th>Adjusted Model</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>b</em></td>
<td><em>e</em>&lt;sup&gt;<em>b</em>&lt;/sup&gt;</td>
<td><em>p</em></td>
<td><em>b</em></td>
<td><em>e</em>&lt;sup&gt;<em>b</em>&lt;/sup&gt;</td>
</tr>
<tr>
<td>MalePt</td>
<td>0.08</td>
<td>1.08</td>
<td>.0016</td>
<td>0.06</td>
<td>1.06</td>
</tr>
<tr>
<td>ln LOS</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>0.92</td>
<td>1.09†</td>
</tr>
</tbody>
</table>

† *e*<sup>*b*ln(1.1)*: a 10% increase in LOS → about a 9% increase in costs

. In the unadjusted model, compared to female patients
  male patients had total costs expected to be about 8% higher

. After conditioning on *ln* LOS, compared to female patients
  male patients had total costs expected to be about 6% higher

. A fraction of the sex effect might be explained by LOS
Example Application: *Logistic Model*

Modeling the effect of patient sex on binary Costs

<table>
<thead>
<tr>
<th></th>
<th>Unadjusted Model</th>
<th></th>
<th>Adjusted Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>UOR</td>
<td>p</td>
</tr>
<tr>
<td>MalePt</td>
<td>0.14</td>
<td>1.15</td>
<td>.0016</td>
</tr>
<tr>
<td>lnLOS</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

† $e^{b*\ln(1.1)}$: a 10% increase in LOS $\rightarrow$ about 54% increased odds of higher costs

. In the unadjusted model, compared to female patients, male patients had about 15% higher odds of 'high costs'

. After conditioning on lnLOS, compared to female patients, male patients had about 35% higher odds of 'high costs'

Change in AOR away from 1.0 might suggest negative confounding
Reconciliation: Is negative confounding a possibility?

An example of negative confounding

. Negative confounding: Multiple effects of X with different signs

. Direct Effect is positive: Education $\rightarrow$ Liberalism

. Indirect Effect is negative: Education $\rightarrow$ $$ $\rightarrow$ Liberalism

. Total Effect = Direct Effect + Indirect Effect

Under negative confounding,

Direct and Indirect Effects tend to cancel one another
Reconciliation: Is negative confounding a possibility?

<table>
<thead>
<tr>
<th>Correlations</th>
<th>MalePt</th>
<th>lnLOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ln{\text{LOS}} )</td>
<td>0.01</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>( p=0.3464 )</td>
<td>( p&lt;.0001 )</td>
</tr>
<tr>
<td>( \ln{\text{Costs}} )</td>
<td>0.04</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>( p=.0002 )</td>
<td>( p&lt;.0001 )</td>
</tr>
<tr>
<td>( \text{binaryCosts} )</td>
<td>0.04</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>( p=.0016 )</td>
<td>( p&lt;.0001 )</td>
</tr>
</tbody>
</table>

Here we see that \( r_{\text{(MalePt lnLOS)}} = 0.01 \) potential for confounding of any type is low
Reconciliation: Is negative confounding a possibility?

Based upon information from the previous two slides, we can create a path diagram:

![Path Diagram]

Summary
- The direct effect (OR = 1.35) is positive
- The indirect effect—a function of \( r=0.01 \) and \( \text{OR}=1.54 \)—is also positive
- No negative confounding
Reconciliation: Underlying cause of discrepant results

**Linear model:** \( y_i = \text{intercept}_{\text{lin}} + x_i b_{\text{lin}} + e_i \)

. The residual variance is estimated and shrinks as the explanatory power of the model increases

**Logistic model:** \( \logit[\Pr(y_i=1 \mid x_i)] = \text{intercept}_{\text{log}} + x_i b_{\text{log}} \)

. The residual variance of the logistic regression model is *fixed*
  . equal to \( \pi^2/3 \) (variance of the standard logistic distribution)
  . this is done for the purpose of model identification
Reconciliation: Underlying cause of discrepant results

Implications of fixing the residual variation in logistic regression

As additional $x$ variables are added to a logistic regression model…

. Residual variance cannot be reduced (it is fixed by assumption)

. Something has to 'give'…

. Implied variation of the outcome increases: is rescaled

. Parameter estimates and ORs are also rescaled

  *All else being equal*…

  . Rescaled parameter estimates move away from zero

  . Rescaled ORs (e.g., AORs) move away from 1.0
Reconciliation: Underlying cause of discrepant results

Implications of fixing the residual variation in logistic regression

Comparing parameters across nested logistic regression models. e.g., AOR versus UOR

<table>
<thead>
<tr>
<th>Operating Condition</th>
<th>Expectation: AOR v UOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>parameter rescaling*</td>
<td>AOR further away from 1.0 than UOR</td>
</tr>
<tr>
<td>negative confounding</td>
<td>AOR further away from 1.0 than UOR</td>
</tr>
<tr>
<td>confounding</td>
<td>AOR closer to 1.0 than UOR</td>
</tr>
<tr>
<td>combination</td>
<td>??? possible counteracting effects</td>
</tr>
</tbody>
</table>

* with nested logistic models, some degree of parameter rescaling is always present
Reconciliation: Underlying cause of discrepant results
…returning to the logistic regression example…

Modeling the effect of patient sex on binaryCosts

<table>
<thead>
<tr>
<th></th>
<th>Unadjusted Model</th>
<th>Adjusted Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>UOR</td>
</tr>
<tr>
<td>MalePt</td>
<td>0.14</td>
<td>1.15</td>
</tr>
<tr>
<td>lnLOS</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

† \( e^{b \ln(1.1)} \): a 10% increase in LOS → about 54% increased odds of higher costs

. The UOR and AOR are on different scales; not directly comparable

. AOR > UOR: ?? neg. confounding, parameter rescaling, or both ??

. For a direct comparison we need a UOR estimate for MalePt that is on the same scale as the corresponding AOR estimate
Reconciliation: A way forward

Estimating a *rescaled* unadjusted effect of MalePt

  . KHB method (Karlson, Holm, and Breen)

---

**Step 1.** Regress $\ln \text{LOS}$ onto MalePt and save residuals: $\ln \text{LOS}_{\text{resid}}$

**Step 2.** Add $\ln \text{LOS}_{\text{resid}}$ as an $x$ variable:

i.e.,

$$\text{logit}[\Pr(\text{binaryCosts}_i = 1)] = \text{intercept} + \text{MalePt}_i b_1 + \ln \text{LOS}_{\text{resid}} i b_2$$

---

The above model estimates an *unadjusted* effect of MalePt on an equivalent scale as the original AOR for MalePt

  . *Method extends to accommodate multiple $x$ variables & covariates*
Reconciliation: A way forward

The logic underlying the KHB method

\[
\text{logit}[\Pr(\text{binaryCosts}_i = 1)] = \text{intercept} + \text{MalePt}_i b_1 + \text{lnLOS}_{\text{resid}} b_2
\]

1a: MalePt and \(\text{lnLOS}_{\text{resid}}\) are uncorrelated

\[\text{i.e., the effect of MalePt is not adjusted by } \text{lnLOS}_{\text{resid}}\]

1b: any shared variation between MalePt and lnLOS

is retained by MalePt, but removed from \(\text{lnLOS}_{\text{resid}}\)

\[\text{i.e., the 'total' effect of MalePt is estimated}\]

2: \(\text{VAR(MalePt + lnLOS}_{\text{resid}}) \approx \text{VAR(MalePt + lnLOS)}\)

\[\text{i.e., scaling of the parameter estimates is equivalent across the rescaled unadjusted and the original adjusted model}\]
Reconciliation: A way forward

Returning to the logistic regression example...

Modeling the effect of patient sex on binaryCosts

<table>
<thead>
<tr>
<th></th>
<th>Rescaled Unadjusted Model (-2LL = 4811.75)</th>
<th></th>
<th>Adjusted Model (-2LL = 4811.75)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>UOR&lt;sub&gt;rescaled&lt;/sub&gt;</td>
<td>p</td>
</tr>
<tr>
<td>MalePt</td>
<td>0.38</td>
<td>1.47</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>lnLOS&lt;sub&gt;resid&lt;/sub&gt;</td>
<td>4.54</td>
<td>1.54&lt;sup&gt;†&lt;/sup&gt;</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>lnLOS</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

<sup>†</sup> $e^{b \cdot \ln(1.1)}$: a 10% increase in LOS $\rightarrow$ about 54% increased odds of higher costs

. The UOR<sub>rescaled</sub> and AOR are comparably scaled

. Evidence of slight confounding: consistent with the observed correlation between MalePt & lnLOS. No negative confounding
Conclusions

Given nested logistic regression models

. Parameter/OR rescaling always occurs as $x$ variables are added

. When all added $x$ variables have near zero effects, then the degree of rescaling will be negligible

. When an added $x$ variable has a substantial effect, then some substantial rescaling will occur
   The degree of rescaling will increase with variance of $x$

I.e., added $x$ variables with large variance and large effects will induce largest levels of rescaling

Again, all types of nested logistic models, not just unadjusted vs. adjusted
Conclusions

Be wary when comparing effects across nested logistic models…

…that appear to suggest negative confounding
  (e.g., AOR is further away from 1.0 than UOR)
  . Probably, you are observing effects of rescaling

…where the OR for a particular \( x \) variable does not appreciably change across nested models
  . Probably you are observing counteracting effects of adjustment by confounders and rescaling

However, if the UOR is substantial, but the AOR is near 1.0,
  then you can attribute the change to adjustment for confounders.
Conclusions

. KHB method is simple to implement

. KHB method seems to do a good job of obtaining rescaled, unadjusted point estimates

. Quality of KHB method standard errors/coverage
   Coverage of rescaled unadjusted $x$ effects was just OK in a limited simulation that I conducted

. If you want to emphasize any tests of rescaled unadjusted effects, the bootstrap should be considered

. Parameter/OR rescaling concerns go beyond logistic regression. Any model with residual variance fixed by assumption will have the same issues.
Resources

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


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