Two postestimation commands for assessing confounding

Zhiqiang Wang
Centre for Chronic Disease
School of Medicine

12th United Kingdom Stata Users Group Meeting, London 2006
Confounding

- A situation in which a measure of the effect of an exposure on risk is distorted because the association with other factor(s) that influence the outcome under study. (Last 1995)
- A bias due to the imbalanced distribution of extraneous risk factor(s) among study groups.
Approaches to handling confounding

- Restriction
- Randomisation
- Matching
- Stratification
- Multiple regression analysis
Practical approach to assessing confounding

- To compare adjusted with crude effect estimates
Multiple regression

- Number of possible effect estimates
  - Number = $2^n$
  (where $n =$ number of confounders)
Regression analysis

- Stepwise (P value)
- Bayesian (BIC)
- Akaike (AIC)
- Deviance (DIC)
- Focussed (FIC)
- Bayesian model averaging
- Frequentist model averaging
- Propensity score (two stages)
help for confall

plot and display all possible effect estimates after estimation

confall [, eform(string) lockterms(varlist) format(fmt) xis(string) xformat(fmt) addaic addbic notable
graph_options]

Description

confall plots and displays effect estimates from models with all possible combinations of potential confounders. The first independent variable is the exposure of interest and all other independent variables are potential confounders.

Options

eform reports the estimated coefficients transformed to odds ratios, hazard ratios, relative risk i.e., exp(b) rather than b. Confidence intervals are similarly transformed.

format(fmt) specifies the display format for presenting numbers in graph and table. format(09.0g) is the default

xformat(fmt) specifies the display format for presenting numbers in x axis. format(09.0g) is the default

lockterms(varlist) specifies variables to be included in all models.

xis() specifies values for x-axis. The default is p representing p values. Alternatives include aic (Akaike Information Criterion), bic (Bayesian Information Criterion), r2 (R2 or Pseudo R2), and n (the number of confounders)

addaic and addbic mark the effect estimate from the model with the minimum AIC and BIC, respectively.

notable suppresses the display of the table.

graph_options refers to options of graph twoway scatter.

Examples

logistic diabetes bmi age sex cholesterol sbp crp smoking drinking
confall, lockterm(age sex) eform(OR)
confall, eform(OR) xline(0.05) yline(1)
confall, eform(OR) yline(1) xis(aic)

Author
Zhichang Wang <z.wang@uq.edu.au>
The University of Queensland

Acknowledgments
This work was supported by the NHMRC Australia (301024).
. stset time, id(id) failure(diabincident==1)

    id: id
  failure event: diabincident == 1
obs. time interval: (time[_n-1], time]
  exit on or before: failure

714  total obs.
    0  exclusions

714  obs. remaining, representing
714  subjects
126  failures in single failure-per-subject data
6725.955  total analysis time at risk, at risk from t = 0
  earliest observed entry t = 0
  last observed exit t = 12.81862
All possible effect estimates: syntax

Exposure variable

Potential confounding variables

Command
stcox CRP BMI Age Sex IGT weight SBP Cholesterol Diastolic Smoking Drinking GGT ACR

Command
confall, lockterms(Age Sex) xis(p) eform(Hazard ratio) addbic addaic /* graphic options

aic bic civ r2 n
```
. cconfall, eform(Hazard ratio) lock(Age Sex)
Please wait........
```

<table>
<thead>
<tr>
<th>Var. adj.</th>
<th>Hazard ratio</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>null</td>
<td>1.571584</td>
<td>1.282775 1.925417</td>
</tr>
<tr>
<td>BMI</td>
<td>1.477466</td>
<td>1.169968 1.865783</td>
</tr>
<tr>
<td>IGT</td>
<td>1.54529</td>
<td>1.252523 1.906736</td>
</tr>
<tr>
<td>weight</td>
<td>1.452142</td>
<td>1.153083 1.828763</td>
</tr>
<tr>
<td>SBP</td>
<td>1.585356</td>
<td>1.277286 1.918545</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>1.584687</td>
<td>1.277 1.967006</td>
</tr>
<tr>
<td>DiastolicBP</td>
<td>1.572189</td>
<td>1.274784 1.914814</td>
</tr>
<tr>
<td>Smoking</td>
<td>1.570635</td>
<td>1.281288 1.924589</td>
</tr>
<tr>
<td>BMI IGT</td>
<td>1.443765</td>
<td>1.138398 1.831045</td>
</tr>
<tr>
<td>BMI weight</td>
<td>1.470657</td>
<td>1.170326 1.870746</td>
</tr>
<tr>
<td>BMI SBP</td>
<td>1.477475</td>
<td>1.169934 1.865756</td>
</tr>
<tr>
<td>BMI Cholesterol</td>
<td>1.475168</td>
<td>1.162041 1.875228</td>
</tr>
<tr>
<td>BMI DiastolicBP</td>
<td>1.475016</td>
<td>1.172519 1.864526</td>
</tr>
<tr>
<td>BMI smoking</td>
<td>1.475407</td>
<td>1.168039 1.863598</td>
</tr>
<tr>
<td>IGT weight</td>
<td>1.425284</td>
<td>1.127072 1.802420</td>
</tr>
<tr>
<td>IGT SBP</td>
<td>1.545171</td>
<td>1.257314 1.907135</td>
</tr>
<tr>
<td>IGT Cholesterol</td>
<td>1.546682</td>
<td>1.238438 1.936616</td>
</tr>
<tr>
<td>IGT DiastolicBP</td>
<td>1.533699</td>
<td>1.247384 1.894845</td>
</tr>
<tr>
<td>IGT smoking</td>
<td>1.542997</td>
<td>1.250843 1.900131</td>
</tr>
<tr>
<td>weight SBP</td>
<td>1.451516</td>
<td>1.157256 1.828053</td>
</tr>
<tr>
<td>weight Cholesterol</td>
<td>1.452017</td>
<td>1.144587 1.842021</td>
</tr>
<tr>
<td>weight DiastolicBP</td>
<td>1.444242</td>
<td>1.147921 1.817053</td>
</tr>
<tr>
<td>weight smoking</td>
<td>1.441014</td>
<td>1.140708 1.827187</td>
</tr>
<tr>
<td>SBP Cholesterol</td>
<td>1.538499</td>
<td>1.257892 1.935269</td>
</tr>
<tr>
<td>SBP DiastolicBP</td>
<td>1.536735</td>
<td>1.257023 1.931508</td>
</tr>
<tr>
<td>SBP smoking</td>
<td>1.561698</td>
<td>1.276611 1.917796</td>
</tr>
<tr>
<td>Cholesterol DiastolicBP</td>
<td>1.576616</td>
<td>1.271765 1.954544</td>
</tr>
<tr>
<td>Cholesterol smoking</td>
<td>1.582786</td>
<td>1.274874 1.965065</td>
</tr>
<tr>
<td>DiastolicBP smoking</td>
<td>1.562138</td>
<td>1.274794 1.941215</td>
</tr>
<tr>
<td>BMI IGT weight</td>
<td>1.445142</td>
<td>1.143515 1.839603</td>
</tr>
<tr>
<td>BMI IGT SBP</td>
<td>1.445835</td>
<td>1.138649 1.831038</td>
</tr>
<tr>
<td>BMI IGT Cholesterol</td>
<td>1.445286</td>
<td>1.132115 1.845023</td>
</tr>
<tr>
<td>BMI IGT DiastolicBP</td>
<td>1.441919</td>
<td>1.137307 1.828118</td>
</tr>
<tr>
<td>BMI IGT smoking</td>
<td>1.443111</td>
<td>1.137493 1.830837</td>
</tr>
<tr>
<td>BMI weight SBP</td>
<td>1.470675</td>
<td>1.170292 1.870848</td>
</tr>
<tr>
<td>BMI weight Cholesterol</td>
<td>1.470696</td>
<td>1.160218 1.880108</td>
</tr>
</tbody>
</table>
Age sex in all models
1034 sets of confounders
Outcome variable: Exposure: CRP
(File Figure 1a.gph saved)

end of do-file
Example 1: C-reactive protein and Diabetes
All effect estimates
Example 1: CRP and Diabetes
All effect estimates
Example 2: No Association and No Confounding
All effect estimates
Example 3: GGT and Diabetes
All effect estimates
help for confnd

plot and display change-in-effect estimates after estimation

```
confnd [ . cform forward lockterms(varlist) format(%fmt) notable graph_options ]
```

Description

`confnd` plots and displays effect estimates using stepwise change-in-estimate method for assessing confounding. The first independent variable is the exposure of interest and all other independent variables are potential confounders.

Options

`cform` reports the estimated coefficients transformed to odds ratios, hazard ratios, relative risk i.e., exp(b) rather than b. Confidence intervals are similarly transformed.

`format(%fmt)` specifies the display format for presenting numbers in graph and table. `format(%9.0g)` is the default.
`forward` specifies the forward method, the backward method is the default.
`lockterms(varlist)` specifies variables to be included in all models.
`notable` suppresses the display of the table.
`graph_options` refers to options of graph twoway scatter.

Examples

```
. logistic diabetes bmi age sex cholesterol sbp cpr smoking drinking
  . confnd, cform
  . confnd, cform(Odds ratio) xline(1)
```

Author

Zhiqiang Wang <zhang@ecs.uq.edu.au>
The University of Queensland

Acknowledgments

This work was supported by the NHMRC Australia (301024).

Also see

On-line: help for confall (if installed)
Change-in-effect estimates: syntax

Command
```
stcox CRP BMI Age Sex IGT weight SBP Cholesterol Diastolic Smoking Drinking GGT ACR
```

Command
```
confd, lockterms(Age Sex) eform(Hazard ratio) forward /* graphic options
```

Default: backward
### Stcox regression

Number of obs = 714

Outcome: CRP

<table>
<thead>
<tr>
<th>Variables added</th>
<th>Hazard Ratio</th>
<th>[95% conf.]</th>
<th>Interval</th>
<th>Change, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial model</td>
<td>1.47</td>
<td>1.28</td>
<td>1.93</td>
<td>7.80</td>
</tr>
<tr>
<td>+weight</td>
<td>1.45</td>
<td>1.18</td>
<td>1.83</td>
<td>1.64</td>
</tr>
<tr>
<td>+ACR</td>
<td>1.40</td>
<td>1.12</td>
<td>1.76</td>
<td>1.26</td>
</tr>
<tr>
<td>+GGT</td>
<td>1.38</td>
<td>1.10</td>
<td>1.74</td>
<td>1.37</td>
</tr>
<tr>
<td>+Cholesterol</td>
<td>1.40</td>
<td>1.11</td>
<td>1.77</td>
<td>1.37</td>
</tr>
<tr>
<td>+BMI</td>
<td>1.42</td>
<td>1.12</td>
<td>1.74</td>
<td>1.26</td>
</tr>
<tr>
<td>+SBP</td>
<td>1.42</td>
<td>1.11</td>
<td>1.80</td>
<td>1.47</td>
</tr>
<tr>
<td>+IGT</td>
<td>1.41</td>
<td>1.10</td>
<td>1.74</td>
<td>0.34</td>
</tr>
<tr>
<td>+DiastolicBP</td>
<td>1.41</td>
<td>1.10</td>
<td>1.79</td>
<td>0.15</td>
</tr>
<tr>
<td>+Smoking</td>
<td>1.40</td>
<td>1.10</td>
<td>1.79</td>
<td>0.19</td>
</tr>
<tr>
<td>+Drinking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Age, Sex in all models

End of do-file

---

Graph (Graph)

Initial model

+Weight
+ACR
+GGT
+Cholesterol
+BMI
+SBP
+IGT
+DiastolicBP
+Smoking
+Drinking

Hazard Ratio
Example 1: CRP and diabetes
change-in-effect estimates: confnd

- Initial model
- +weight
- +ACR
- +GGT
- +Cholesterol
- +BMI
- +SBP
- +IGT
- +DiastolicBP
- +Smoking
- +Drinking

Hazard Ratio
### Change-in-estimate

**stcox regression.**

**Number of obs = 714**

**Outcome:**

**Exposure: CRP**

<table>
<thead>
<tr>
<th>Variables added</th>
<th>Hazard Ratio</th>
<th>[95% Conf. Interval]</th>
<th>Change, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial model</td>
<td>1.57</td>
<td>1.28</td>
<td>1.93</td>
</tr>
<tr>
<td>+weight</td>
<td>1.45</td>
<td>1.15</td>
<td>1.83</td>
</tr>
<tr>
<td>+ACR</td>
<td>1.40</td>
<td>1.12</td>
<td>1.76</td>
</tr>
<tr>
<td>+GGT</td>
<td>1.38</td>
<td>1.10</td>
<td>1.74</td>
</tr>
<tr>
<td>+Cholesterol</td>
<td>1.40</td>
<td>1.11</td>
<td>1.77</td>
</tr>
<tr>
<td>+BMI</td>
<td>1.42</td>
<td>1.12</td>
<td>1.80</td>
</tr>
<tr>
<td>+SBP</td>
<td>1.42</td>
<td>1.11</td>
<td>1.80</td>
</tr>
<tr>
<td>+IGT</td>
<td>1.41</td>
<td>1.11</td>
<td>1.80</td>
</tr>
<tr>
<td>+DiastolicBP</td>
<td>1.41</td>
<td>1.10</td>
<td>1.80</td>
</tr>
<tr>
<td>+Smoking</td>
<td>1.40</td>
<td>1.10</td>
<td>1.79</td>
</tr>
<tr>
<td>+Drinking</td>
<td>1.40</td>
<td>1.10</td>
<td>1.79</td>
</tr>
</tbody>
</table>

*Age Sex in all models*

*end of do-file*
```stata
. confnd, for format(%6.2f)

Change-in-estimate
stcox regression.
number of obs = 714

Outcome: Exposure: CRP

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coef.</th>
<th>[95% Conf. Interval]</th>
<th>Change, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude</td>
<td>0.49</td>
<td>0.30</td>
<td>0.68</td>
</tr>
<tr>
<td>+ACR</td>
<td>0.36</td>
<td>0.17</td>
<td>0.56</td>
</tr>
<tr>
<td>+weight</td>
<td>0.34</td>
<td>0.12</td>
<td>0.55</td>
</tr>
<tr>
<td>+Sex</td>
<td>0.35</td>
<td>0.13</td>
<td>0.57</td>
</tr>
<tr>
<td>+GGT</td>
<td>0.33</td>
<td>0.11</td>
<td>0.56</td>
</tr>
<tr>
<td>+Cholesterol</td>
<td>0.35</td>
<td>0.12</td>
<td>0.58</td>
</tr>
<tr>
<td>+Age</td>
<td>0.35</td>
<td>0.10</td>
<td>0.57</td>
</tr>
<tr>
<td>+BMI</td>
<td>0.35</td>
<td>0.11</td>
<td>0.59</td>
</tr>
<tr>
<td>+SBP</td>
<td>0.35</td>
<td>0.11</td>
<td>0.59</td>
</tr>
<tr>
<td>+IGT</td>
<td>0.34</td>
<td>0.10</td>
<td>0.59</td>
</tr>
<tr>
<td>+DiastolicBP</td>
<td>0.34</td>
<td>0.10</td>
<td>0.59</td>
</tr>
<tr>
<td>+Smoking</td>
<td>0.34</td>
<td>0.10</td>
<td>0.58</td>
</tr>
<tr>
<td>+Drinking</td>
<td>0.34</td>
<td>0.09</td>
<td>0.58</td>
</tr>
</tbody>
</table>
```

Example 2: No Association and No Confounding
Change-in-effect estimates: confnd

Variables added

Initial model
+BMI
+IGT
+ACR
+DiastolicBP
+Cholesterol
+SBP
+CRP
+weight
+Smoking
+Drinking
+GGT

Hazard Ratio
Example 3: GGT and Diabetes
Change-in-effect estimates: confnd

- Initial model
- +Cholesterol
- +IGT
- +ACR
- +weight
- +BMI
- +CRP
- +SBP
- +DiastolicBP
- +Smoking
- +Drinking

Hazard Ratio

0.80 1.00 1.20 1.40 1.60
### STCOX Regression Results

**Outcome:** GGT  
**Exposure:** GGT  
**Number of obs:** 714

<table>
<thead>
<tr>
<th>Variables Added</th>
<th>Hazard Ratio</th>
<th>[95% Conf.]</th>
<th>Interval</th>
<th>Change, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial model</td>
<td>1.32</td>
<td>1.12</td>
<td>1.56</td>
<td></td>
</tr>
<tr>
<td>+Cholesterol</td>
<td>1.13</td>
<td>0.94</td>
<td>1.37</td>
<td>14.03</td>
</tr>
<tr>
<td>+IGT</td>
<td>1.04</td>
<td>0.86</td>
<td>1.27</td>
<td>7.97</td>
</tr>
<tr>
<td>+ACR</td>
<td>1.00</td>
<td>0.82</td>
<td>1.23</td>
<td>4.17</td>
</tr>
<tr>
<td>+weight</td>
<td>1.03</td>
<td>0.83</td>
<td>1.27</td>
<td>2.62</td>
</tr>
<tr>
<td>+BMI</td>
<td>0.99</td>
<td>0.79</td>
<td>1.23</td>
<td>3.71</td>
</tr>
<tr>
<td>+CRP</td>
<td>0.96</td>
<td>0.77</td>
<td>1.20</td>
<td>2.85</td>
</tr>
<tr>
<td>+SBP</td>
<td>0.97</td>
<td>0.78</td>
<td>1.22</td>
<td>1.30</td>
</tr>
<tr>
<td>+DiastolicBP</td>
<td>0.97</td>
<td>0.77</td>
<td>1.21</td>
<td>0.70</td>
</tr>
<tr>
<td>+Smoking</td>
<td>0.96</td>
<td>0.77</td>
<td>1.20</td>
<td>0.43</td>
</tr>
<tr>
<td>+Drinking</td>
<td>0.97</td>
<td>0.77</td>
<td>1.22</td>
<td>0.57</td>
</tr>
</tbody>
</table>

**Age** and **Sex** included in all models.
Example 3: GGT and Diabetes
Change-in-effect estimates: confnd, backward
Conclusions

- Tools are useful for identifying the presence of confounding
- Not a substitute for careful incorporation of available knowledge to select confounding factors in the design stage, or careful data analysis.
Acknowledgements

This work was supported by the National Health and Medical Research Council (NHMRC) of Australia (301024, 320860).