Two postestimation commands for assessing confounding

Zhiqiang Wang

Centre for Chronic Disease

School of Medicine

THE UNIVERSITY OF QUEENSLAND

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)

Viewer (#1) [help confall]

Back | Refresh | Search | Help! | Contents | What's New | News

Command: help confall

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(%9.0g)** is the default

xformat(*%fmt*) specifies the display format for presenting numbers in x axis. **format(%9.0g)** 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)
```

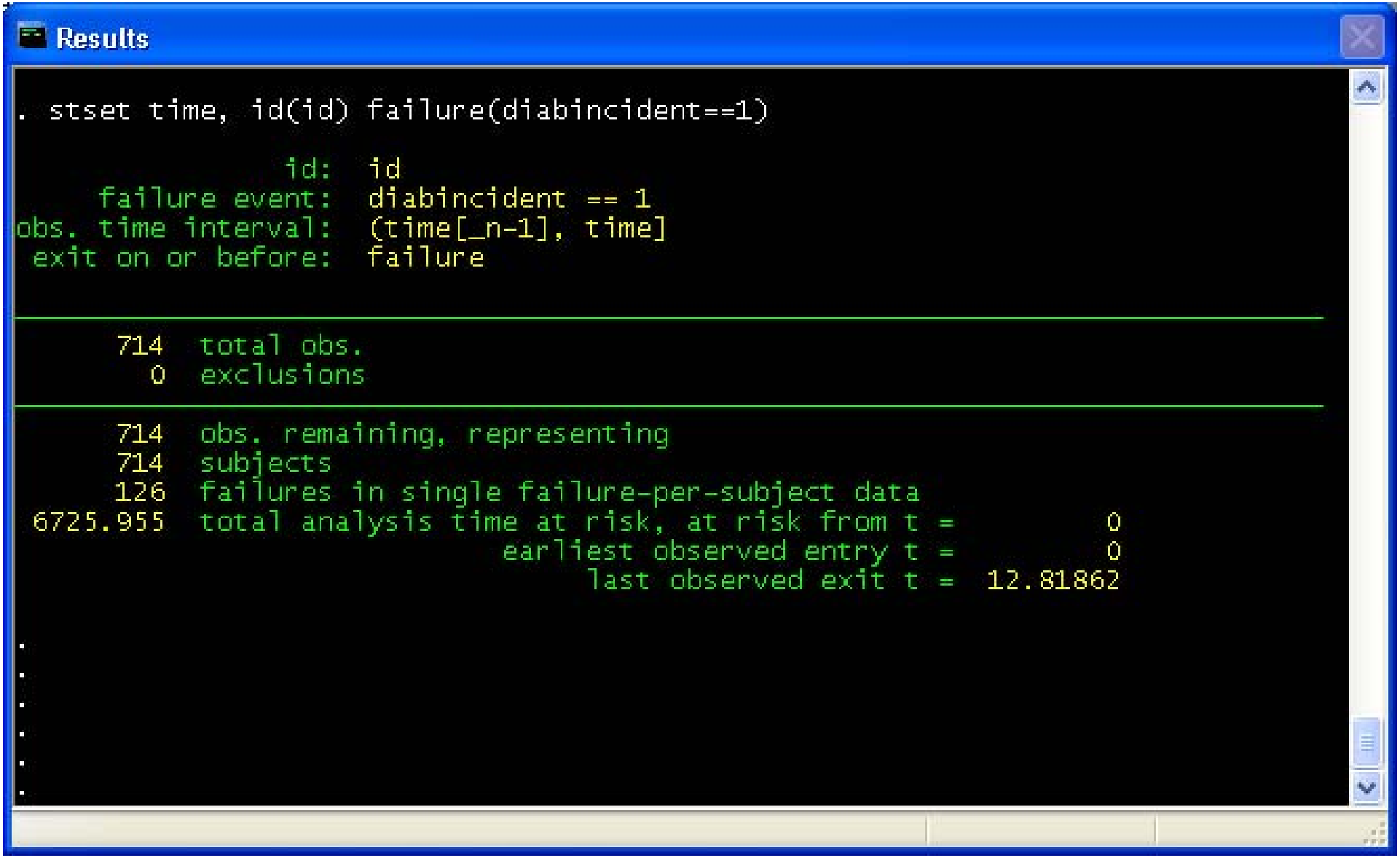
Author

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

Acknowledgments

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

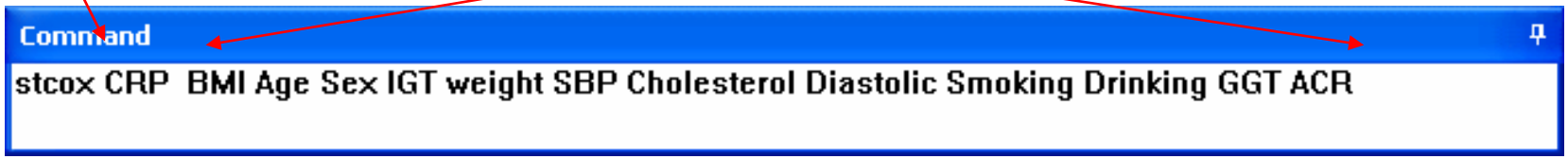
65 1 Read Ovr Block Sync Rec Caps


Results
. 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
.
.
.
.
.
.

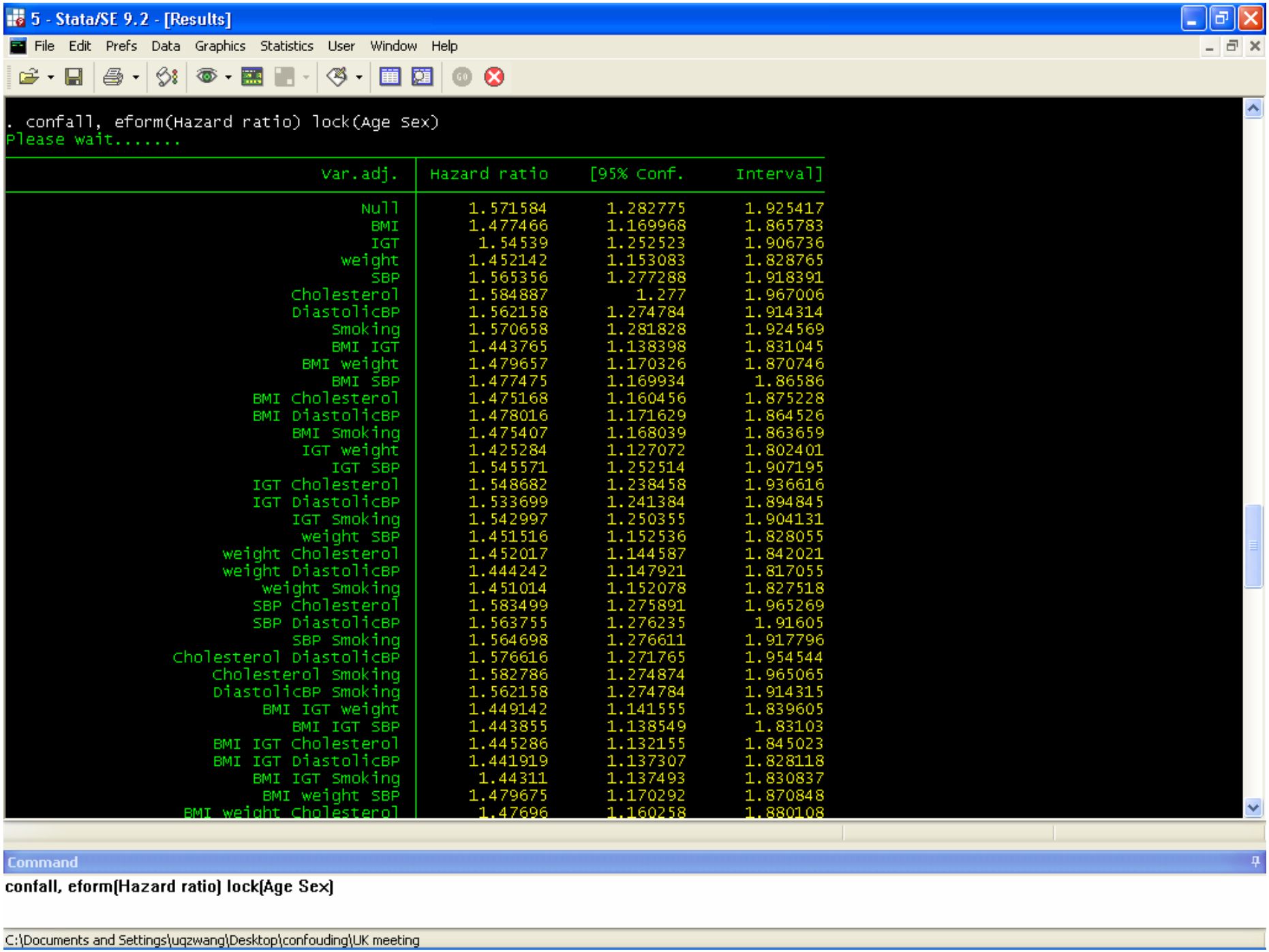
All possible effect estimates: syntax

Exposure variable

Potential confounding variables



aic bic civ r2 n

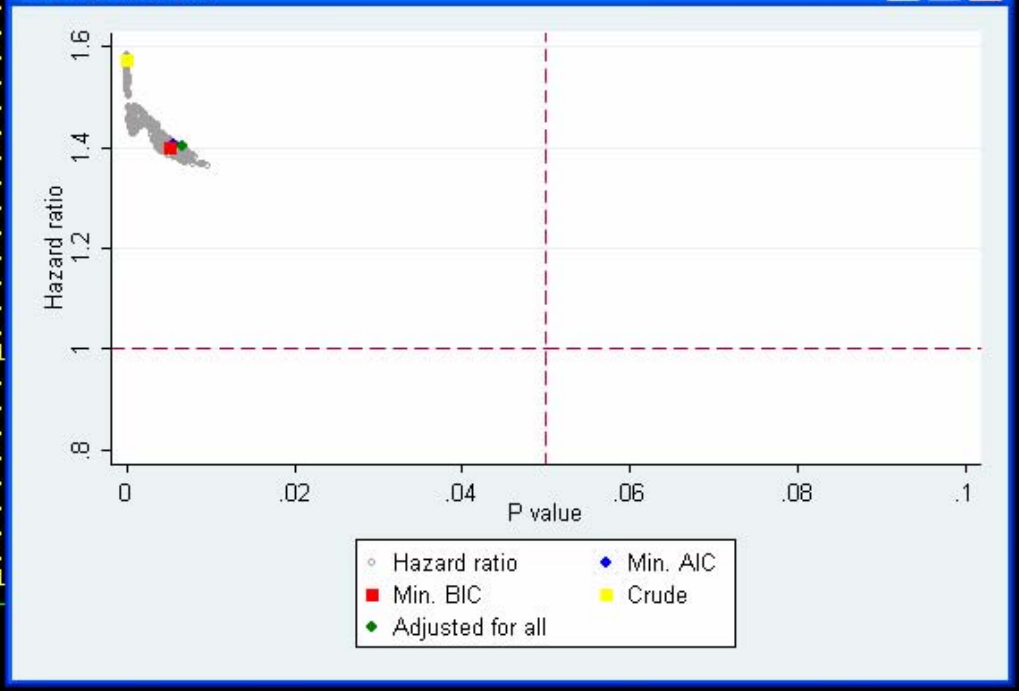

5 - Stata/SE 9.2 - [Results]
File Edit Prefs Data Graphics Statistics User Window Help
. confall, eform(Hazard ratio) lock(Age Sex)
Please wait.....
Var.adj. | Hazard ratio [95% Conf. Interval]
Null 1.571584 1.282775 1.925417
BMI 1.477466 1.169968 1.865783
IGT 1.54539 1.252523 1.906736
weight 1.452142 1.153083 1.828765
SBP 1.565356 1.277288 1.918391
Cholesterol 1.584887 1.277 1.967006
DiastolicBP 1.562158 1.274784 1.914314
Smoking 1.570658 1.281828 1.924569
BMI IGT 1.443765 1.138398 1.831045
BMI weight 1.479657 1.170326 1.870746
BMI SBP 1.477475 1.169934 1.86586
BMI Cholesterol 1.475168 1.160456 1.875228
BMI DiastolicBP 1.478016 1.171629 1.864526
BMI Smoking 1.475407 1.168039 1.863659
IGT weight 1.425284 1.127072 1.802401
IGT SBP 1.545571 1.252514 1.907195
IGT Cholesterol 1.548682 1.238458 1.936616
IGT DiastolicBP 1.533699 1.241384 1.894845
IGT Smoking 1.542997 1.250355 1.904131
weight SBP 1.451516 1.152536 1.828055
weight Cholesterol 1.452017 1.144587 1.842021
weight DiastolicBP 1.444242 1.147921 1.817055
weight Smoking 1.451014 1.152078 1.827518
SBP Cholesterol 1.583499 1.275891 1.965269
SBP DiastolicBP 1.563755 1.276235 1.91605
SBP Smoking 1.564698 1.276611 1.917796
Cholesterol DiastolicBP 1.576616 1.271765 1.954544
Cholesterol Smoking 1.582786 1.274874 1.965065
DiastolicBP Smoking 1.562158 1.274784 1.914315
BMI IGT weight 1.449142 1.141555 1.839605
BMI IGT SBP 1.443855 1.138549 1.83103
BMI IGT Cholesterol 1.445286 1.132155 1.845023
BMI IGT DiastolicBP 1.441919 1.137307 1.828118
BMI IGT Smoking 1.44311 1.137493 1.830837
BMI weight SBP 1.479675 1.170292 1.870848
BMI weight Cholesterol 1.47696 1.160258 1.880108
Command
confall, eform(Hazard ratio) lock(Age Sex)
C:\Documents and Settings\uqzwang\Desktop\confouding\UK meeting

Stata/SE 9.2 - [Results]

File Edit Prefs Data Graphics Statistics User Window Help

```
BMI IGT SBP DiastolicBP Smoking Drinking      1.37994     1.086834     1.752093
BMI IGT Cholesterol DiastolicBP Smoking       1.40699     1.104992     1.791525
BMI weight SBP Cholesterol DiastolicBP S     1.463452     1.146653     1.867778
BMI weight SBP Cholesterol DiastolicBP S     1.406241     1.108334     1.784222
BMI weight SBP Cholesterol DiastolicBP S     1.412383     1.112417     1.782236
BMI weight SBP Cholesterol DiastolicBP D            1
BMI weight SBP Cholesterol Smoking Drink         1.
BMI weight SBP DiastolicBP Smoking Drink         1.
BMI weight Cholesterol DiastolicBP Smoki         1.
BMI SBP Cholesterol DiastolicBP Smoking          1.
IGT weight SBP Cholesterol DiastolicBP S         1.
IGT weight SBP Cholesterol DiastolicBP S         1.
IGT weight SBP Cholesterol DiastolicBP S         1.
IGT weight SBP Cholesterol DiastolicBP D         1.
IGT weight SBP Cholesterol Smoking Drink         1.
IGT weight SBP DiastolicBP Smoking Drink         1.
IGT weight Cholesterol DiastolicBP Smoki         1.
IGT SBP Cholesterol DiastolicBP Smoking          1.
weight SBP Cholesterol DiastolicBP Smoki         1.
BMI IGT weight SBP Cholesterol Diastolic         1.
BMI IGT weight SBP Cholesterol Diastolic          1
BMI IGT weight SBP Cholesterol Diastolic         1.
BMI IGT weight SBP Cholesterol Diastolic         1.
BMI IGT weight SBP Cholesterol Smoking D         1.
BMI IGT weight SBP DiastolicBP Smoking D         1.
BMI IGT weight Cholesterol DiastolicBP S         1.
BMI IGT SBP Cholesterol DiastolicBP Smok         1.
BMI weight SBP Cholesterol DiastolicBP S         1.
IGT weight SBP Cholesterol DiastolicBP S         1.
BMI IGT weight SBP Cholesterol Diastolic          1

 Age Sex in all models
1024 sets of confounders
Outcome variale:       Exposure: CRP
(file Figure1a.gph saved)

.
end of do-file

.
```

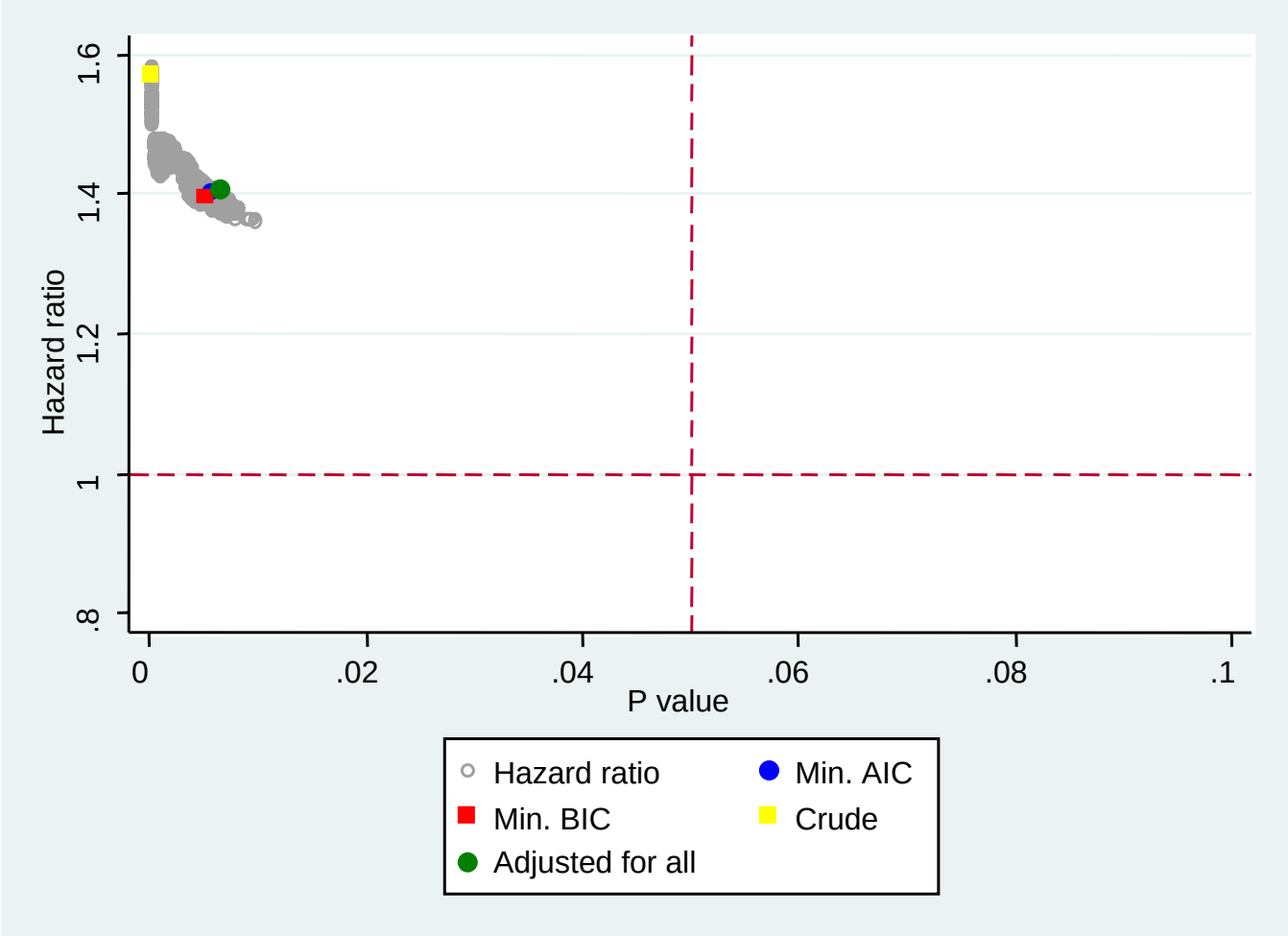
Graph (Graph)



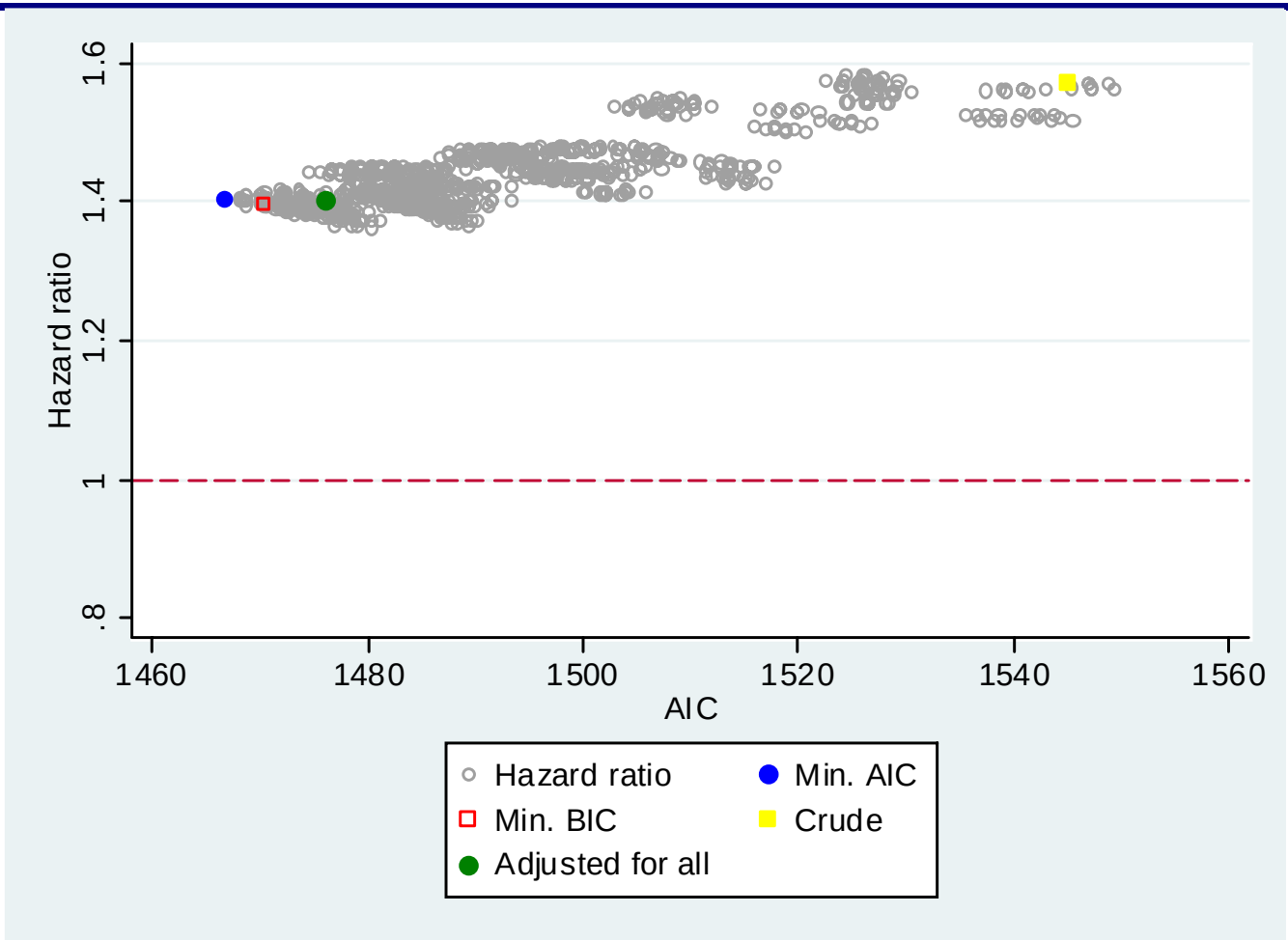
Command

C:\Documents and Settings\uqzwang\Desktop\confouding\UK meeting

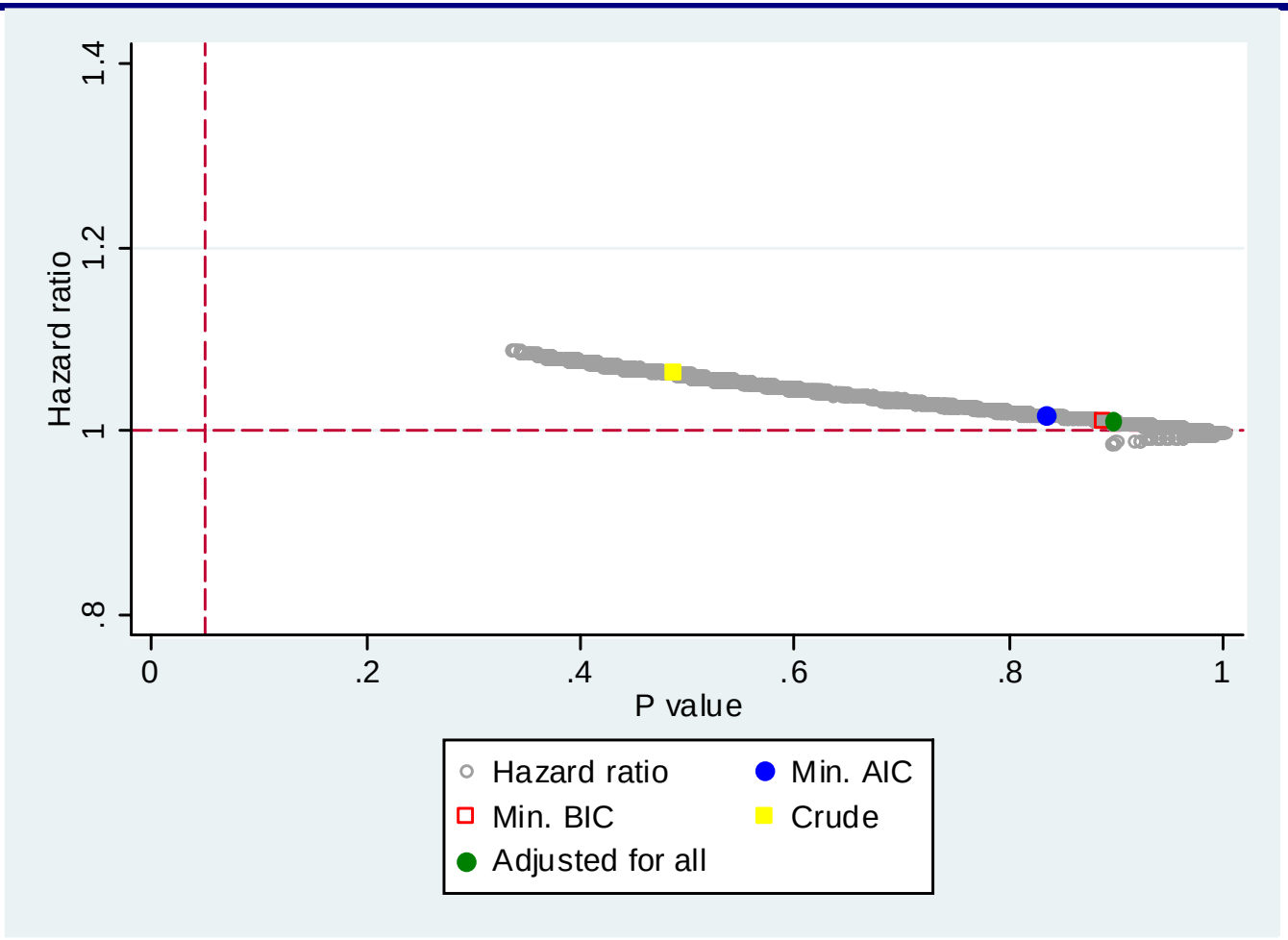
Example 1: C-reactive protein and Diabetes
All effect estimates



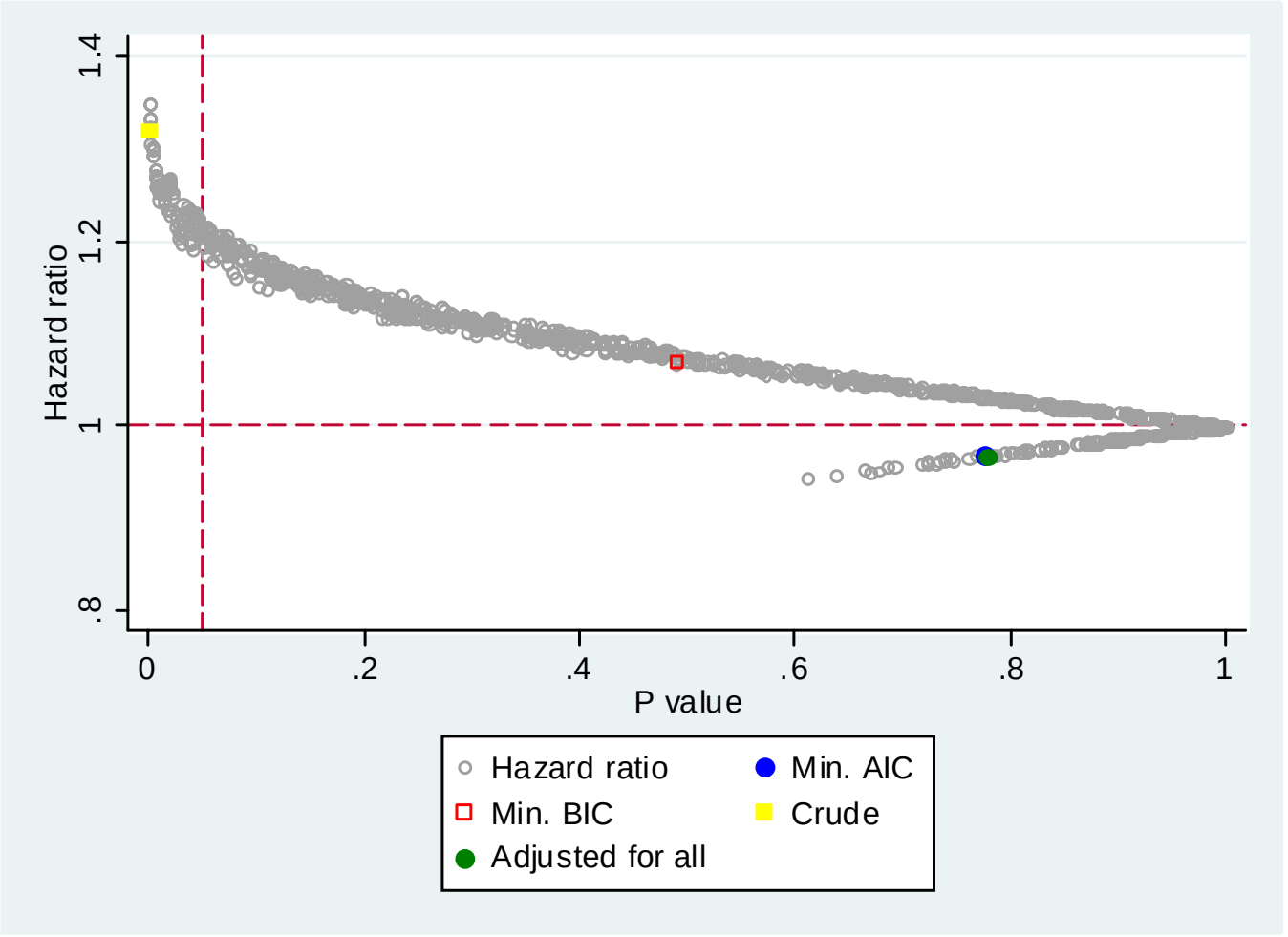
Example 1: CRP and Diabetes
All effect estimates



Example 2: No Association and No Confounding
All effect estimates



Example 3: GGT and Diabetes
All effect estimates



Viewer (#2) [help confnd]

Back | Refresh | Search | Help! | Contents | What's New | News

Command: help confnd

---

help for **confnd**

---

plot and display change-in-effect estimates after estimation

confnd [, **eform 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

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(%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 crp smoking drinking
. confnd, eform
. confnd, eform(Odds ratio) xline(1)
```

Author

Zhiqiang Wang <zwang@ccs.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
confnd, lockterms(Age Sex) eform(Hazard ratio)  forward /* graphic options
```

Default: backward

Stata/SE 9.2

File Edit Prefs Data Graphics Statistics User Window Help

Close

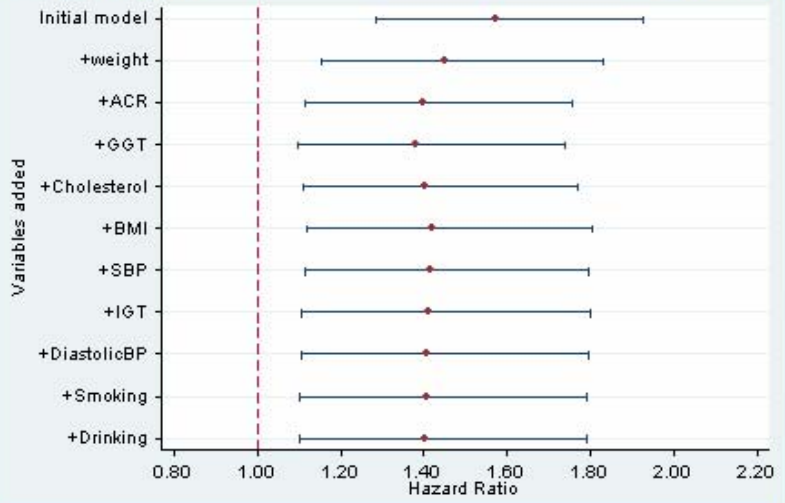
Results

```
stcox regression.                          Outcome:
number of obs = 714                        Exposure: CRP
```

Variables added	Hazard Ratio	[95% Conf.	Interval]	Change, %
Initial model	1.57	1.28	1.93	
+weight	1.45	1.15	1.83	7.60
+ACR	1.40	1.12	1.76	3.64
+GGT	1.38	1.10	1.74	1.26
+Cholesterol	1.40	1.11	1.77	1.37
+BMI	1.42	1.12	1.80	1.47
+SBP	1.42	1.11	1.80	0.40
+IGT	1.41	1.11	1.80	0.34
+DiastolicBP	1.41	1.10	1.80	0.21
+Smoking	1.41	1.10	1.79	0.15
+Drinking	1.40	1.10	1.79	0.19

```
 Age Sex in all models

.
end of do-file
```

Graph (Graph)

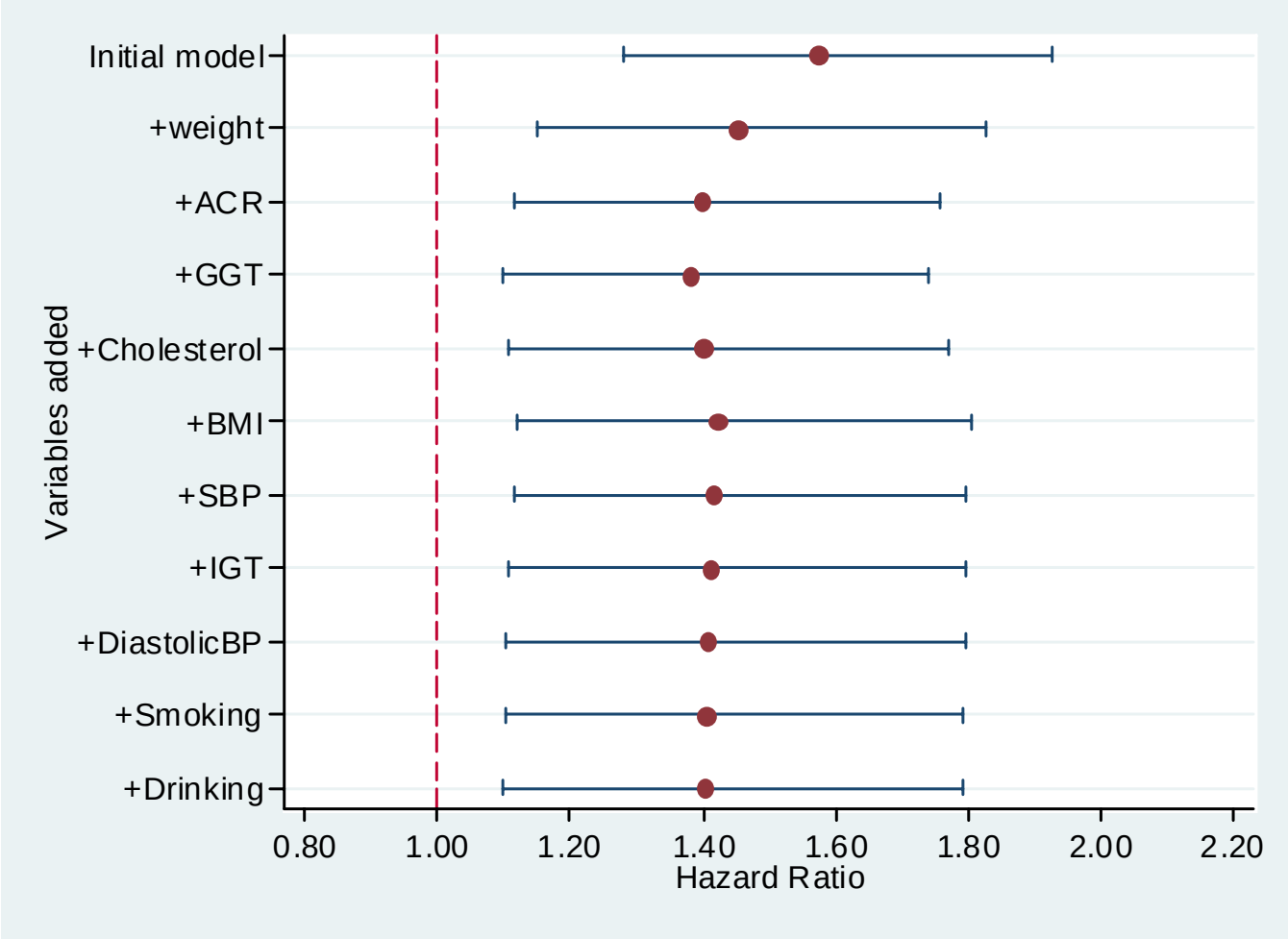


Command

C:\Documents and Settings\uqzwang\Desktop\confouding\UK meeting

Example 1: CRP and diabetes

change-in-effect estimates: confnd



Stata/SE 9.2

File Edit Prefs Data Graphics Statistics User Window Help

Results

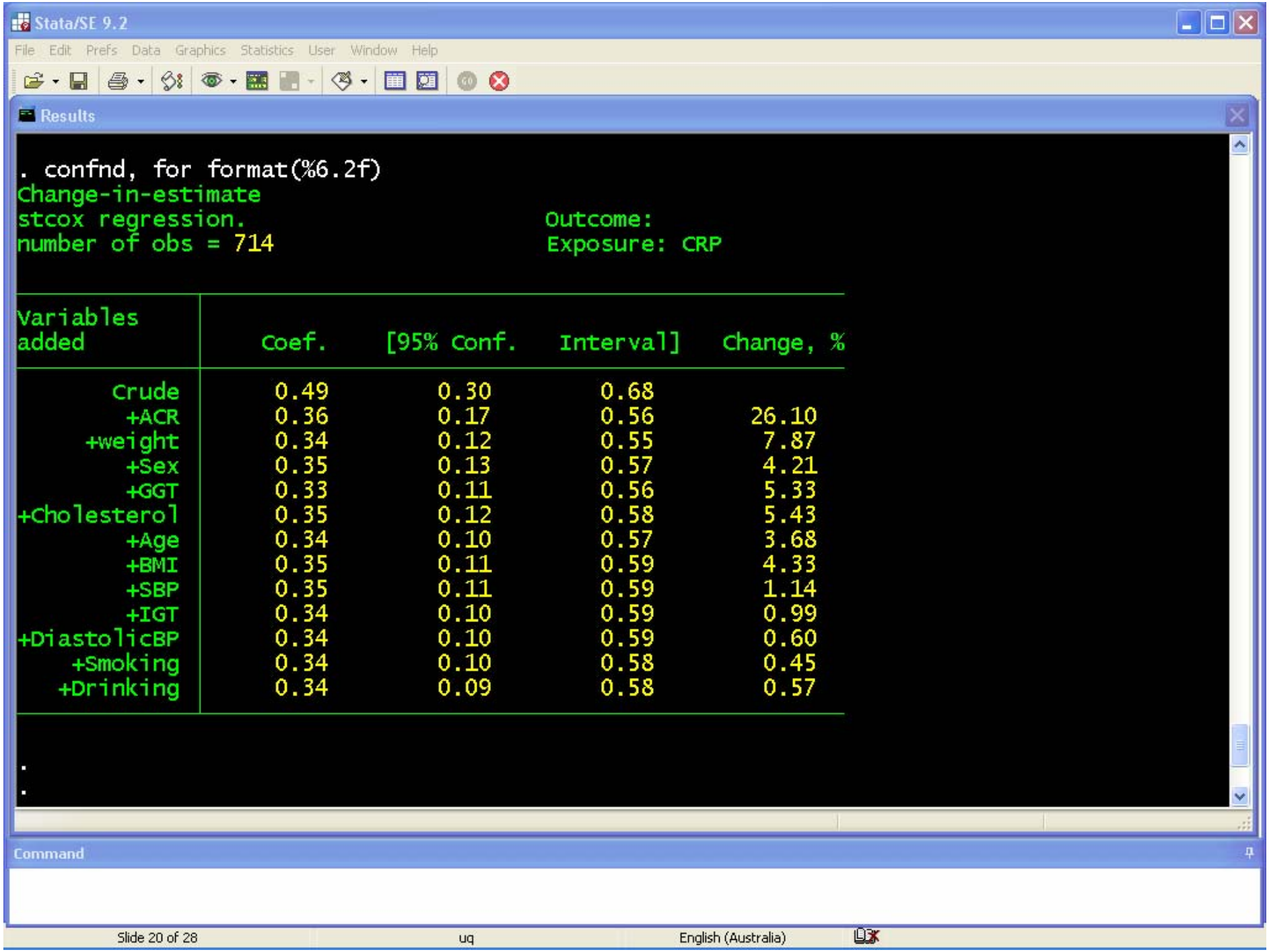
```
Change-in-estimate
stcox regression.                       Outcome:
number of obs = 714                     Exposure: CRP
```

Variables added	Hazard Ratio	[95% Conf.	Interval]	Change, %
Initial model	1.57	1.28	1.93	
+weight	1.45	1.15	1.83	7.60
+ACR	1.40	1.12	1.76	3.64
+GGT	1.38	1.10	1.74	1.26
+Cholesterol	1.40	1.11	1.77	1.37
+BMI	1.42	1.12	1.80	1.47
+SBP	1.42	1.11	1.80	0.40
+IGT	1.41	1.11	1.80	0.34
+DiastolicBP	1.41	1.10	1.80	0.21
+Smoking	1.41	1.10	1.79	0.15
+Drinking	1.40	1.10	1.79	0.19

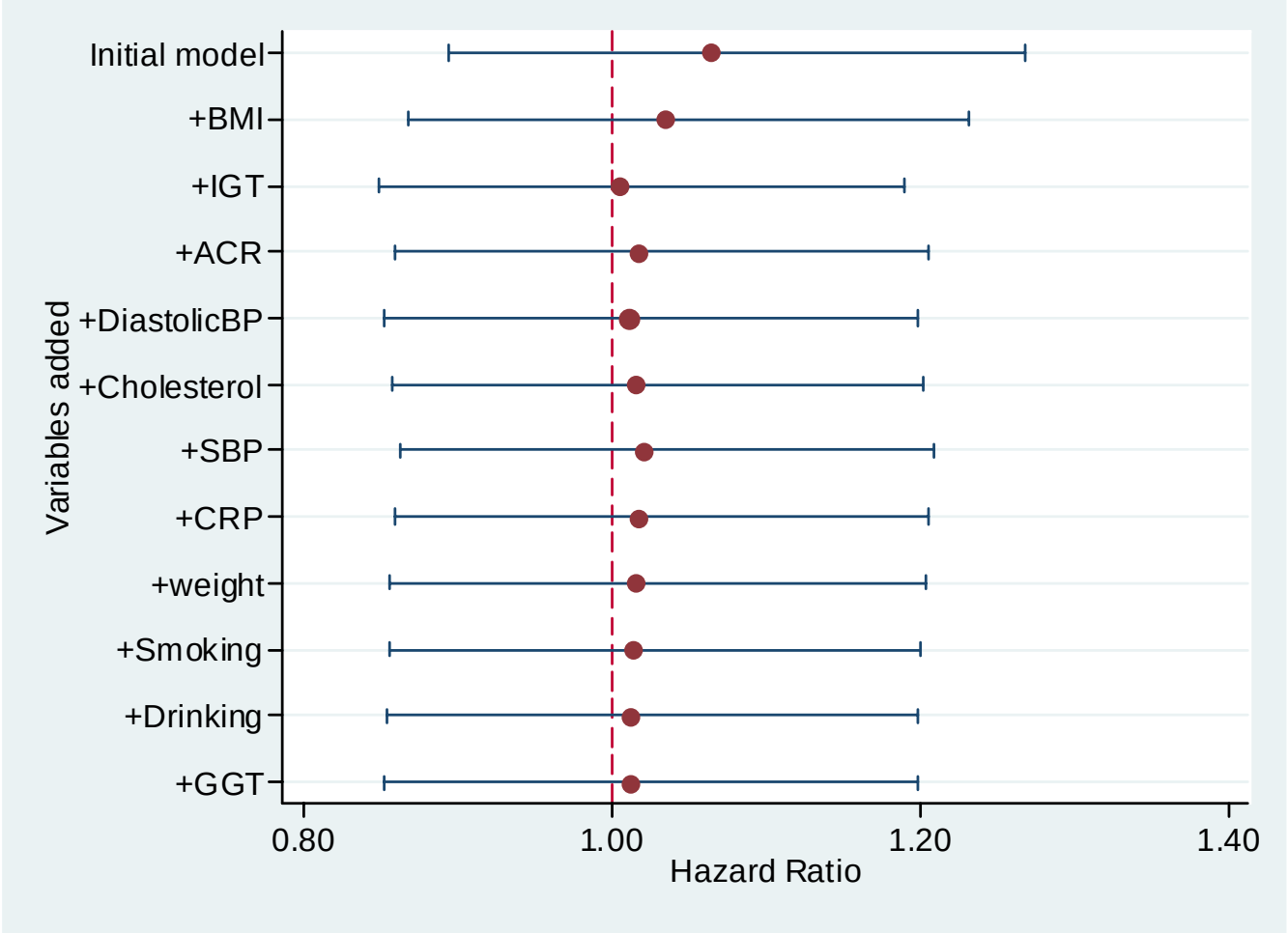
```
 Age Sex in all models

.
end of do-file

.
.
```

Command

Slide 20 of 28 uq English (Australia)

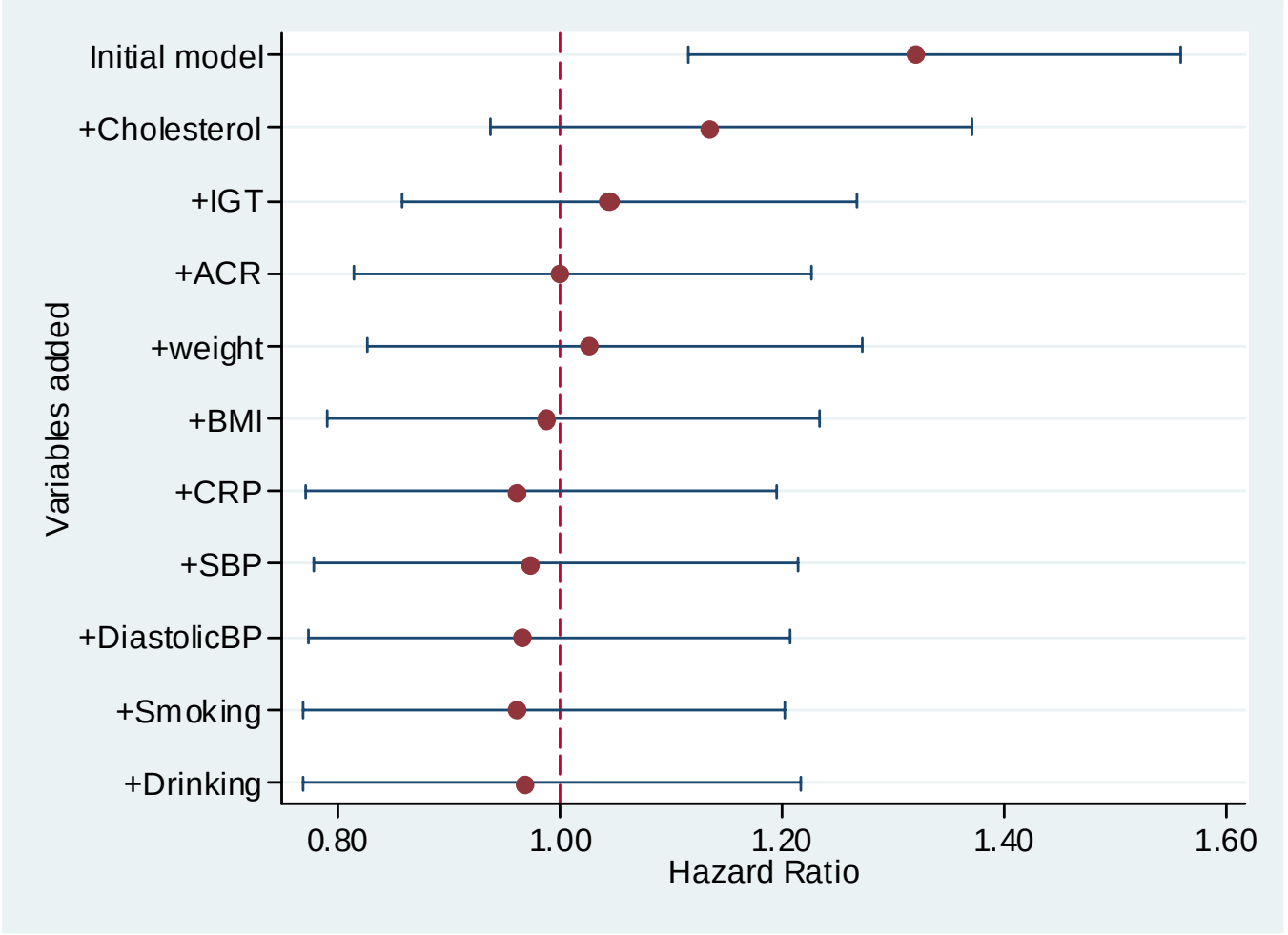

Stata/SE 9.2
File Edit Prefs Data Graphics Statistics User Window Help
Results
. confnd, for format(%6.2f)
Change-in-estimate
stcox regression.
Outcome:
number of obs = 714
Exposure: CRP
Variables added | Coef. [95% Conf. Interval] Change, %
Crude 0.49 0.30 0.68
+ACR 0.36 0.17 0.56 26.10
+weight 0.34 0.12 0.55 7.87
+Sex 0.35 0.13 0.57 4.21
+GGT 0.33 0.11 0.56 5.33
+Cholesterol 0.35 0.12 0.58 5.43
+Age 0.34 0.10 0.57 3.68
+BMI 0.35 0.11 0.59 4.33
+SBP 0.35 0.11 0.59 1.14
+IGT 0.34 0.10 0.59 0.99
+DiastolicBP 0.34 0.10 0.59 0.60
+Smoking 0.34 0.10 0.58 0.45
+Drinking 0.34 0.09 0.58 0.57
Command
Slide 20 of 28
uq
English (Australia)

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



Example 3: GGT and Diabetes

Change-in-effect estimates: confnd



Stata/SE 9.2

File Edit Prefs Data Graphics Statistics User Window Help

Results

Change-in-estimate
stcox regression. Outcome:
number of obs = 714 Exposure: GGT

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

Age Sex in all models

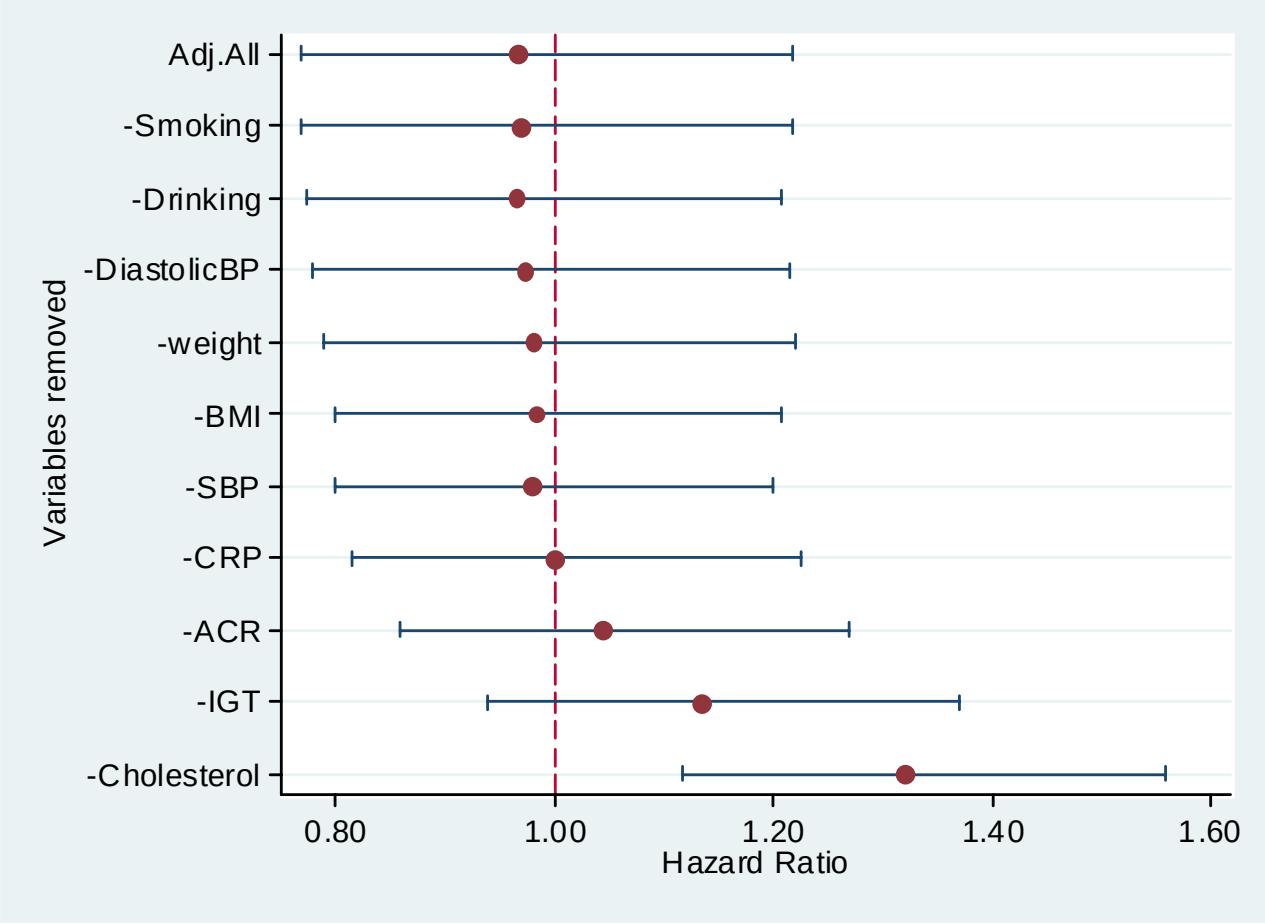
.
end of do-file

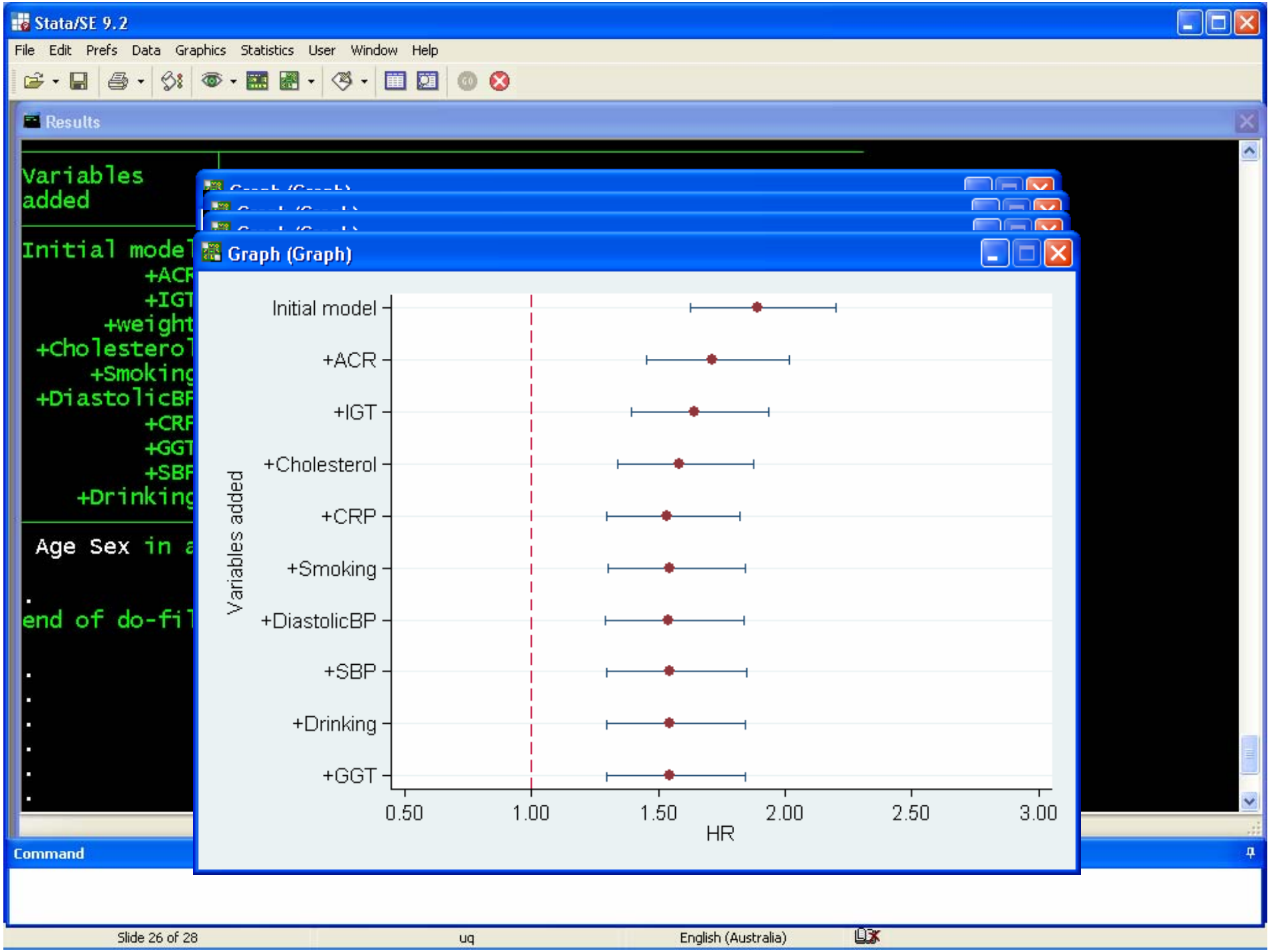
.
.

Command

Example 3: GGT and Diabetes

Change-in-effect estimates: confnd, backward




Stata/SE 9.2
File Edit Prefs Data Graphics Statistics User Window Help
Results
Variables added
Initial model
+ACR
+IGT
+weight
+Cholesterol
+Smoking
+DiastolicBP
+CRP
+GGT
+SBP
+Drinking
Age Sex in a
end of do-fil
Graph (Graph)
Initial model
+ACR
+IGT
+Cholesterol
+CRP
+Smoking
+DiastolicBP
+SBP
+Drinking
+GGT
Variables added
0.50
1.00
1.50
2.00
2.50
3.00
HR
Command
Slide 26 of 28
uq
English (Australia)

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).