

Stata: Visualizing Regression Models Using **coefplot**

Partially based on Ben Jann's June 2014 presentation at the
12th German Stata Users Group meeting in Hamburg, Germany:

“A new command for plotting regression coefficients and other estimates”

http://www.stata.com/meeting/germany14/abstracts/materials/de14_jann.pdf

and on other materials from coefplot website:

<http://repec.sowi.unibe.ch/stata/coefplot/>

Dawn Koffman
Office of Population Research (OPR)
Princeton University
May 9, 2017



Workshop Outline

Motivation

Introduction

coefplot command

- Basic usage
 - single model
 - multiple models
 - subgraphs
- Labels
- Confidence intervals

Motivation

- regression results are often presented in tables

	diabetes
female	1.066 (0.102)
age	1.059*** (0.004)
bmi	1.077*** (0.009)
region==NE	1.081 (0.156)
region==MW	1.091 (0.148)
region==S	1.324** (0.170)

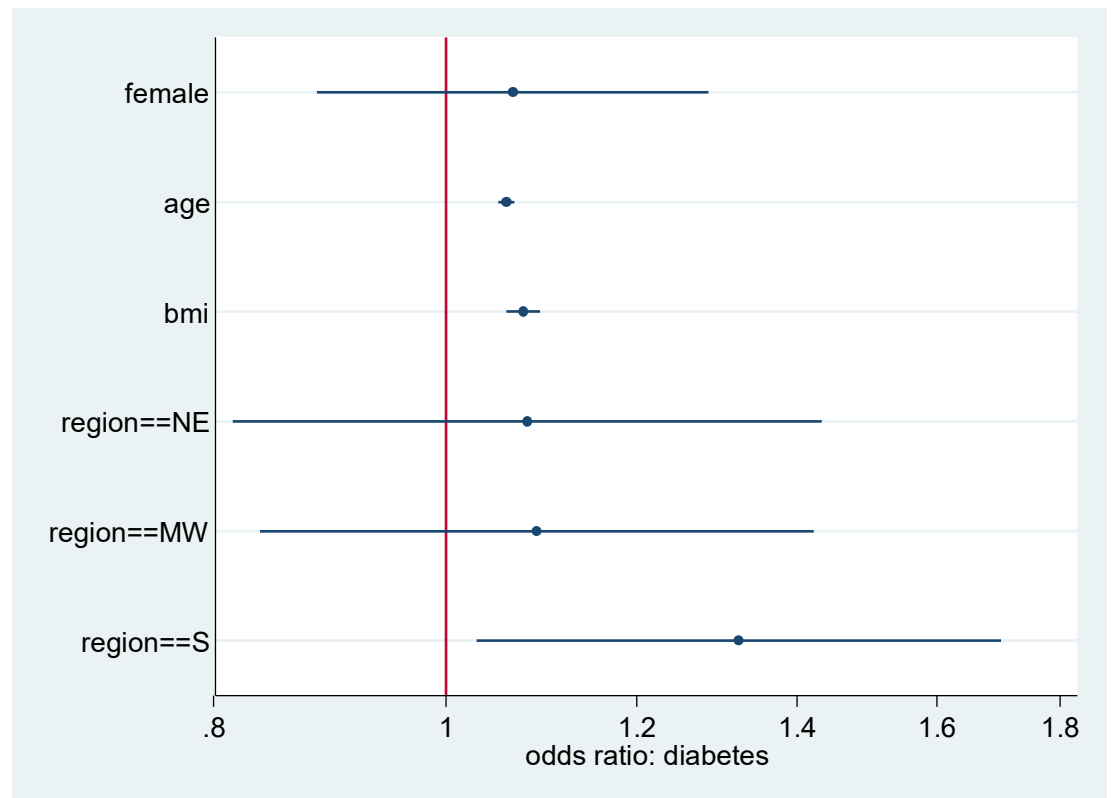
*** p<0.01, ** p<0.05, * p<0.1

Motivation

- regression results are often presented in tables
- **however, displaying results graphically can be much more effective: easier to see and remember patterns and trends**

	diabetes
female	1.066 (0.102)
age	1.059*** (0.004)
bmi	1.077*** (0.009)
region==NE	1.081 (0.156)
region==MW	1.091 (0.148)
region==S	1.324** (0.170)

*** p<0.01, ** p<0.05, * p<0.1

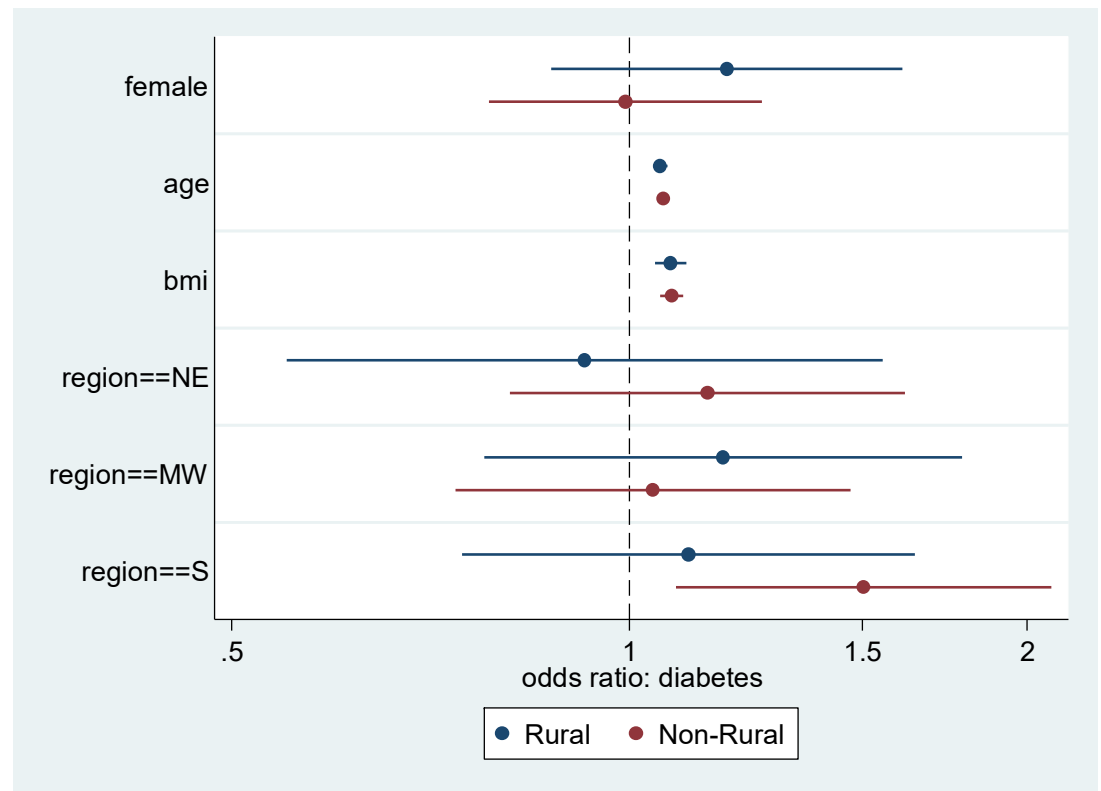


Motivation

- regression results are often presented in tables
- however, displaying results graphically can be much more effective: easier to see and remember patterns and trends

OUTCOME	diabetes	
SAMPLE	Rural	Non-Rural
female	1.185	0.993
	(0.185)	(0.121)
age	1.055***	1.061***
	(0.007)	(0.005)
bmi	1.075***	1.077***
	(0.015)	(0.011)
region==NE	0.925	1.145
	(0.245)	(0.201)
region==MW	1.178	1.042
	(0.250)	(0.183)
region==S	1.108	1.504**
	(0.223)	(0.251)

*** p<0.01, ** p<0.05, * p<0.1



Motivation

For many more examples, see:

Kastellec, Jonathan P.,
Eduardo L. Leoni (2007).
Using Graphs Instead of Tables in Political Science.

Perspectives on Politics
5(4): 755-771.

Table 8
Pekkanen, Nyblade and Krauss (2006),
table 1: Logit analysis of electoral
incentives and LDP post allocation
(1996–2003)

Variable	Model 1	Model 2
<i>Block 1: MP Type</i>		
Zombie	0.18 (.22)	0.27 (0.22)
SMD Only	-0.19 (0.22)	-0.19 (0.24)
PR Only	-0.39 (0.18)**	—
Costa Rican in PR	-0.09 (0.29)	—
<i>Block 2: Electoral Strength</i>		
Vote share margin	—	0.005 (0.004)
Margin Squared	—	—
<i>Block 3: Misc Controls</i>		
Urban-Rural Index	0.04 (0.08)	0.04 (0.09)
No Factional Membership	-0.86 (0.26)***	-0.98 (0.31)***
Legal Professional	0.39 (0.29)	-0.36 (0.30)
<i>Seniority</i>		
1 st Term	-3.76 (0.36)***	-3.66 (0.37)***
2 nd Term	-1.61 (0.19)***	-1.59 (0.21)***
4 th Term	-0.34 (0.19)***	-0.45 (0.21)***
5 th Term	-1.17 (0.22)***	-1.24 (0.24)***
6 th Term	-1.15 (0.22)***	-1.04 (0.24)***
7 th Term	-1.52 (0.25)***	-1.83 (0.29)***
8 th Term	-1.66 (0.28)***	-1.82 (0.32)***
9 th Term	-1.34 (0.32)***	-1.21 (0.33)***
10 th Term	-2.89 (0.48)***	-2.77 (0.49)***
11 th Term	-1.88 (0.43)***	-1.34 (0.46)***
12 th Term	-1.08 (0.41)***	-0.94 (0.49)**
Constant	.020 (.20)	0.13 (0.26)
Log-likelihood	-917.24	-764.77
N	1895	1574

Notes: Dependent Variables: 1 if MP holds a post of minister, vice minister, PARC, or HoR Committee Chair.

Base categories: SMD dual-listed, 3rd term. Excluded observations: senior MPs that held no post (> 12 terms, PR-Only MPs in Model 2).

Figure 7
Using parallel dot plots with error bars to
present two regression models.

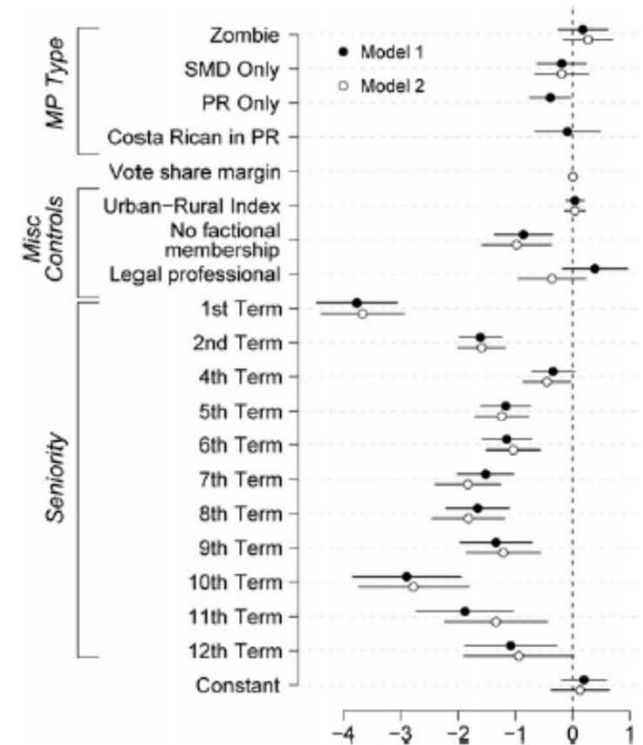


Table 1 from Pekkanen et al. 2006 displays two logistic regression models that examine the allocation of posts in the LDP party in Japan. We turn the table into a graph, and present the two models by plotting parallel lines for each of them grouped by coefficients. We differentiate the models by plotting different symbols for the point estimates: filled (black) circles for Model 1 and empty (white) circles for Model 2.

Tables vs Graphs

tables

useful for seeing individual, precise values

graphs

allow reader to easily:

- see patterns within a set of regression results
- make comparisons

particularly useful when presenting results:

- from several models
- using different samples

coefplot Introduction

Stata command for graphing results of Stata estimation commands

user-written

- author: Ben Jann, University of Bern

default behavior

- plots markers for coefficients and horizontal spikes for confidence intervals

features

- results from multiple models can be displayed on a single graph
- results from multiple models can be displayed on multiple subgraphs
- labels can be applied to coefficients, groups of coefficients, and subgraphs
- sub-headings can be inserted to structure the display
- confidence intervals can show multiple levels

coefplot Resources

coefplot website: <http://repec.sowi.unibe.ch/stata/coefplot/>

Stata Journal article by Ben Jann

Volume 14, Number 4

Plotting regression coefficients and other estimates

<http://www.stata-journal.com/sjpdf.html?articlenum=gr0059>

Working Paper by Ben Jann

Current version February 9, 2017, First version August 25, 2013

Plotting regression coefficients and other estimates in Stata

<http://repec.sowi.unibe.ch/files/wp1/jann-2013-coefplot.pdf>

Citing coefplot

Author request:

“Thanks for citing coefplot in your work in one of the following ways:

Jann, Ben (2014). Plotting regression coefficients and other estimates. The Stata Journal 14(4): 708-737.

Jann, Ben (2014). Plotting regression coefficients and other estimates in Stata. University of Bern Social Sciences Working Papers No. 1. Available from <http://ideas.repec.org/p/bss/wpaper/1.html>.

Jann, Ben (2013). COEFPLOT: Stata module to plot regression coefficients and other results. Available from <http://ideas.repec.org/c/boc/bocode/s457686.html>.”

Install User-Written coefplot Command

In Stata, run command:

```
ssc install coefplot, replace
```

To view the help file:

```
help coefplot
```

A Single Model

```
webuse nhanes2, clear
tab region, gen(reg)
```

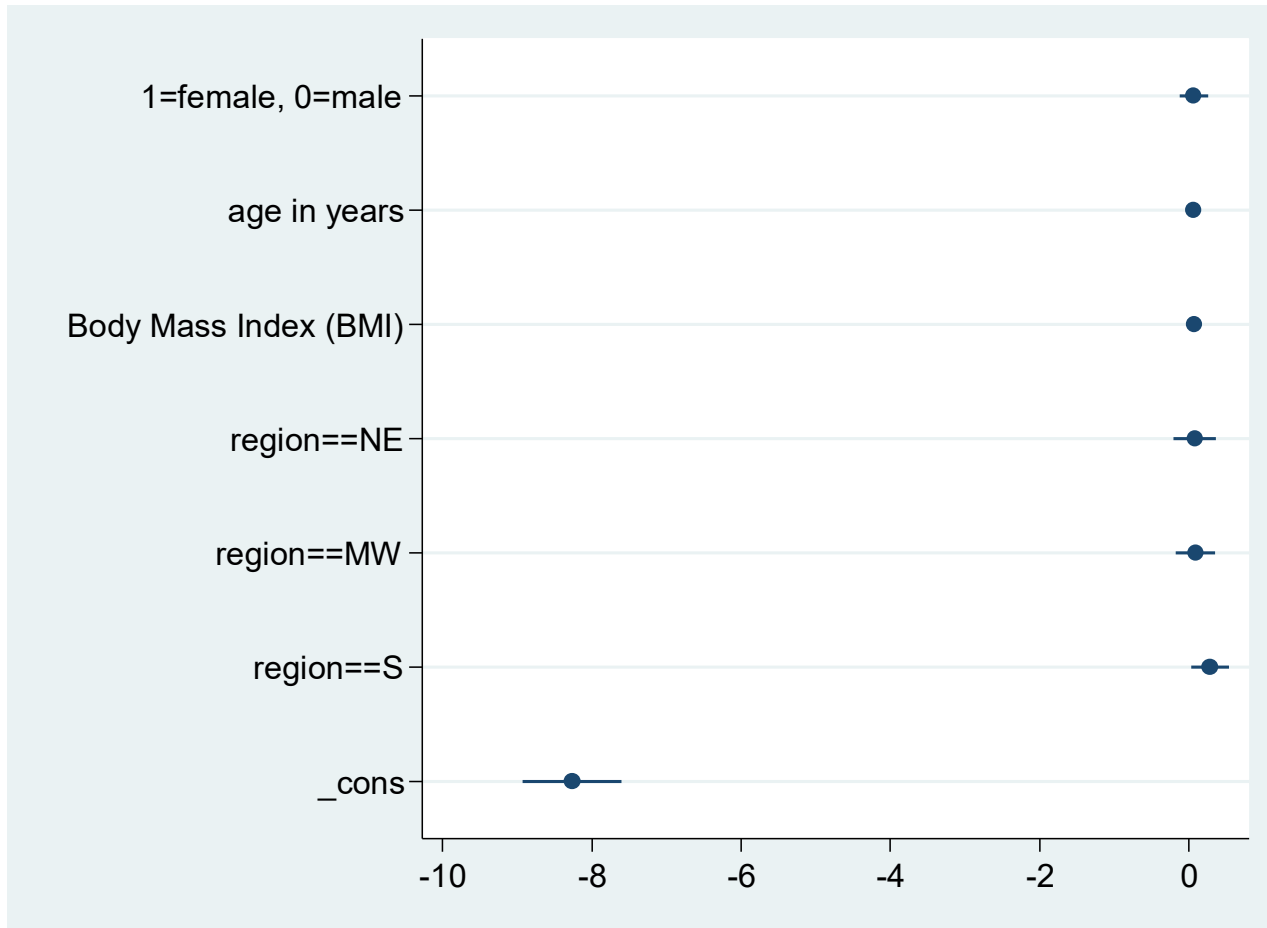
```
logit diabetes female age bmi reg1 reg2 reg3 reg4, or
```

diabetes	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
female	1.0658	.1020453	0.67	0.506	.8834407 1.285803
age	1.059328	.0040456	15.09	0.000	1.051428 1.067287
bmi	1.076815	.0088939	8.96	0.000	1.059524 1.094389
reg1	1.080705	.1555003	0.54	0.590	.8151347 1.432797
reg2	1.090745	.1475434	0.64	0.521	.8367247 1.421883
reg3	1.323504	.1696527	2.19	0.029	1.029472 1.701516
reg4	1	(omitted)			
_cons	.0002576	.0000868	-24.51	0.000	.000133 .0004987

```
estimates store fem_age_bmi_reg
```

Basic Usage: Plotting a Single Model

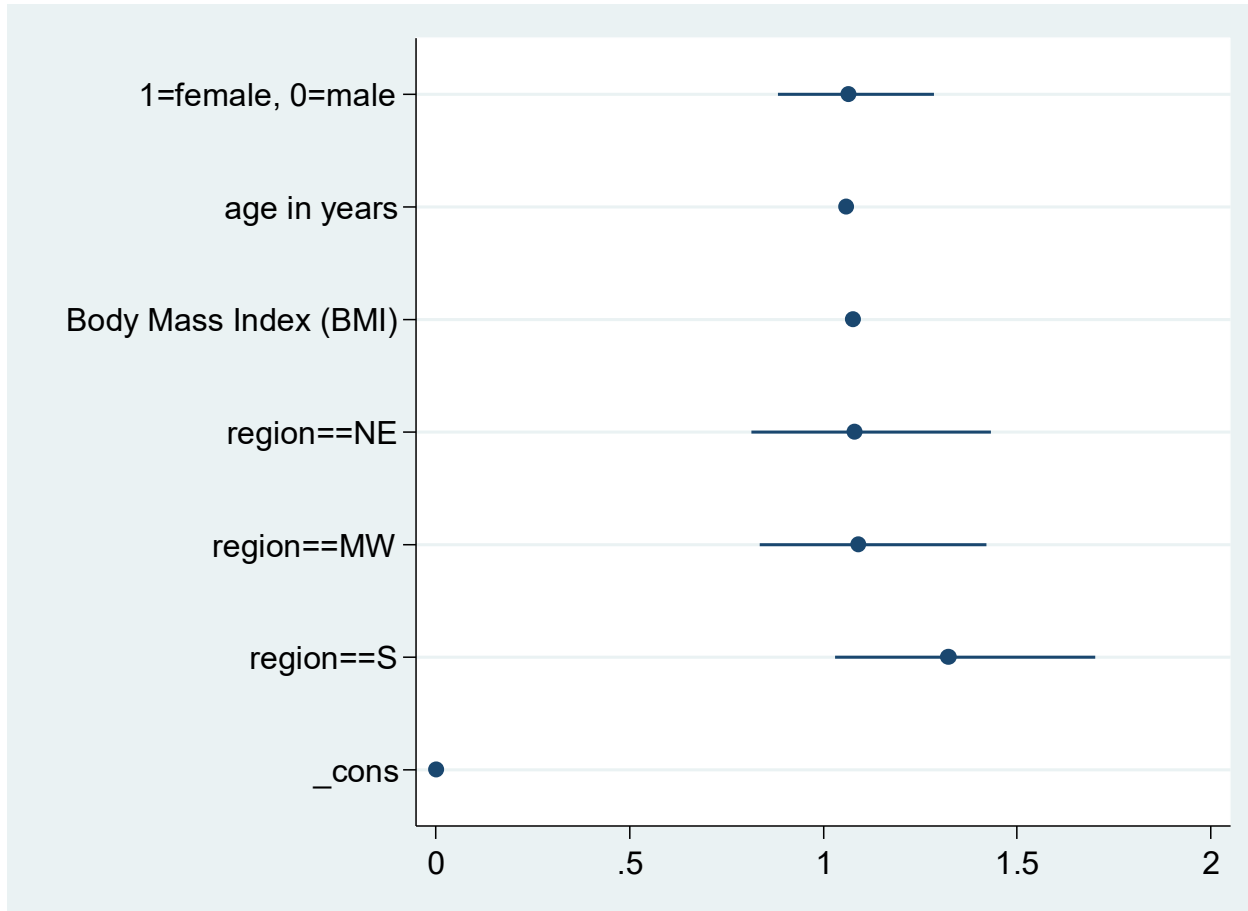
```
coefplot fem_age_bmi_reg
```



clearly, not
odds ratios!

Basic Usage: Plotting a Single Model

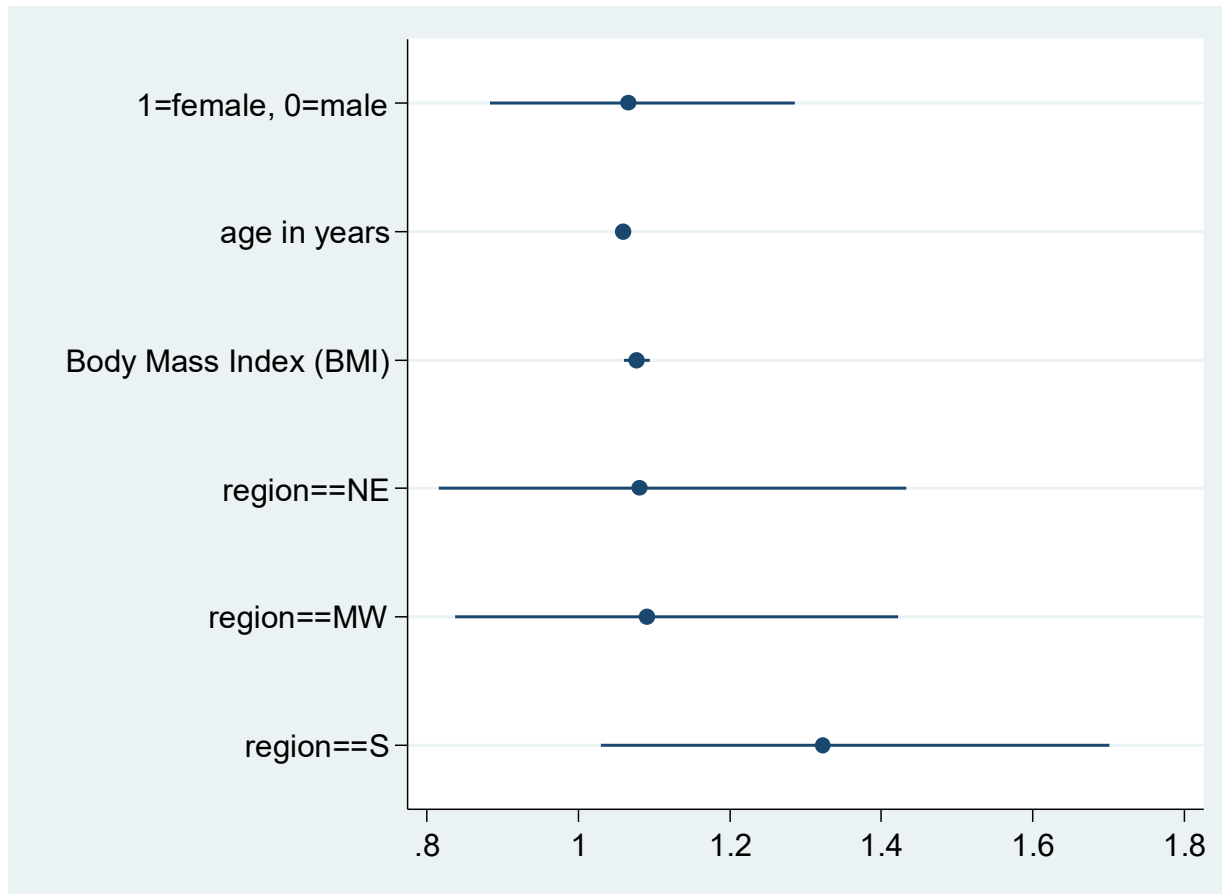
```
coefplot fem_age_bmi_reg, eform
```



odds ratio's,
but don't want
to display constant

Basic Usage: Plotting a Single Model

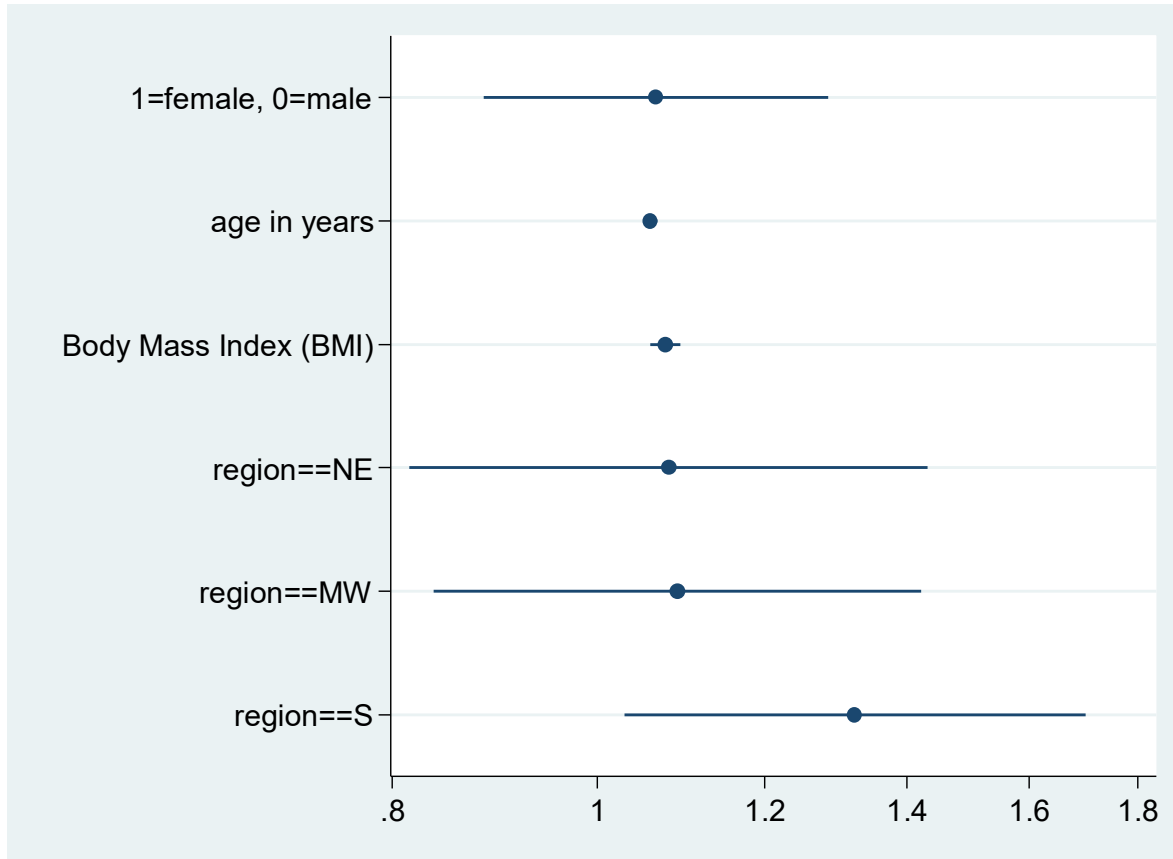
```
coefplot fem_age_bmi_reg, eform drop(_cons)
```



much better, but use linear or log scale to display odds ratio's ??

Basic Usage: Plotting a Single Model

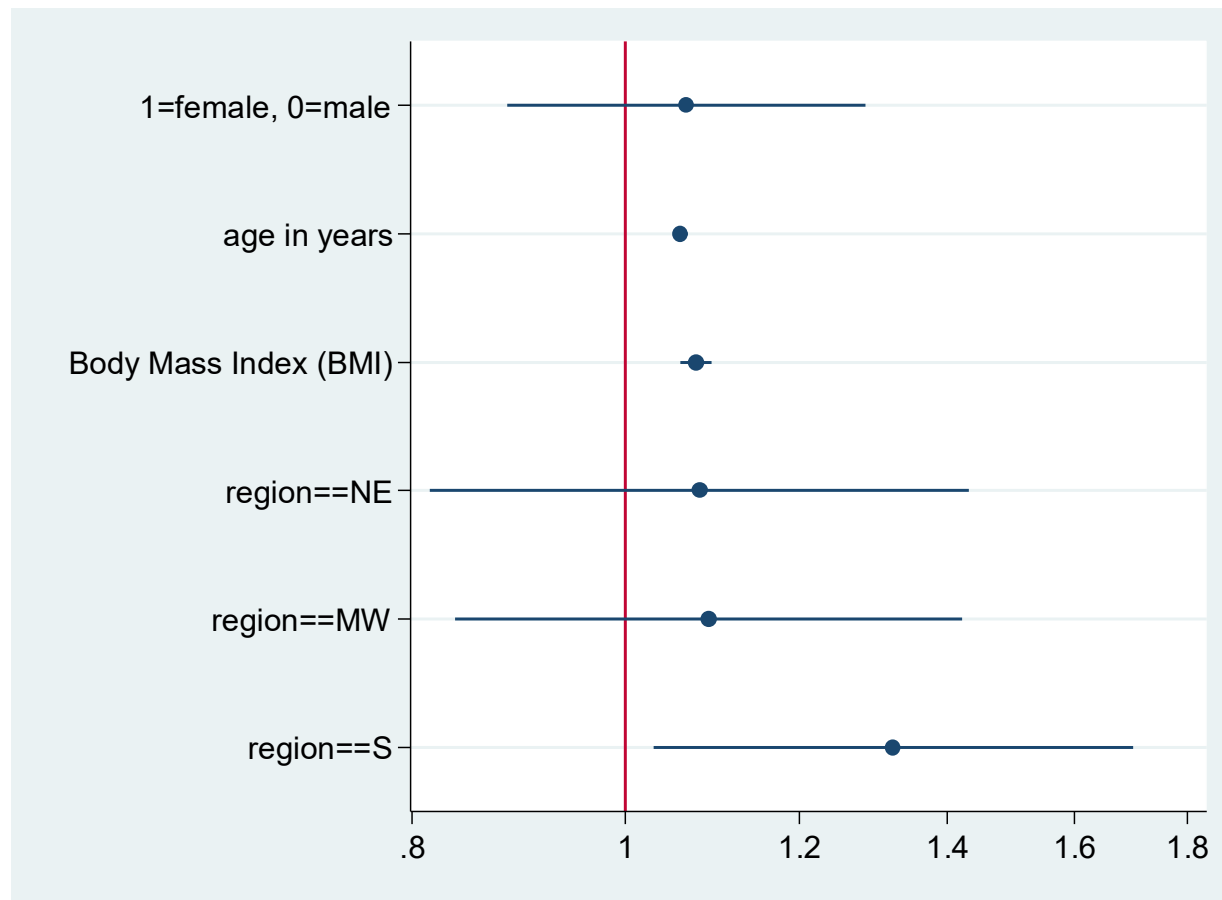
```
coefplot fem_age_bmi_reg, eform drop(_cons) xscale(log)
```



add a line
at $x == 1$?

Basic Usage: Plotting a Single Model

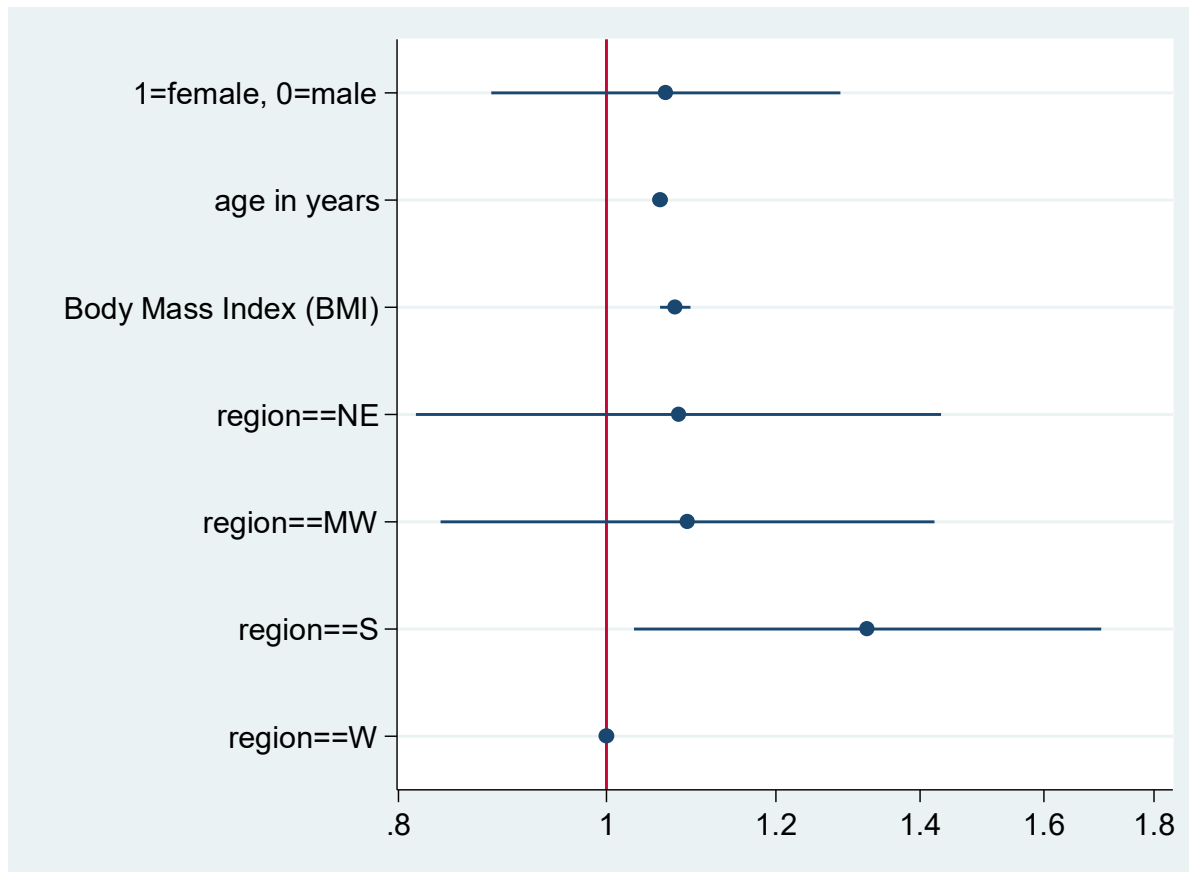
```
#delimit ;  
coefplot fem_age_bmi_reg,  
eform drop(_cons) xscale(log) xline(1)  
;
```



show omitted
odds ratio?

Basic Usage: Plotting a Single Model

```
#delimit ;  
coefplot fem_age_bmi_reg,  
eform drop(_cons) xscale(log) xline(1) omitted  
;
```

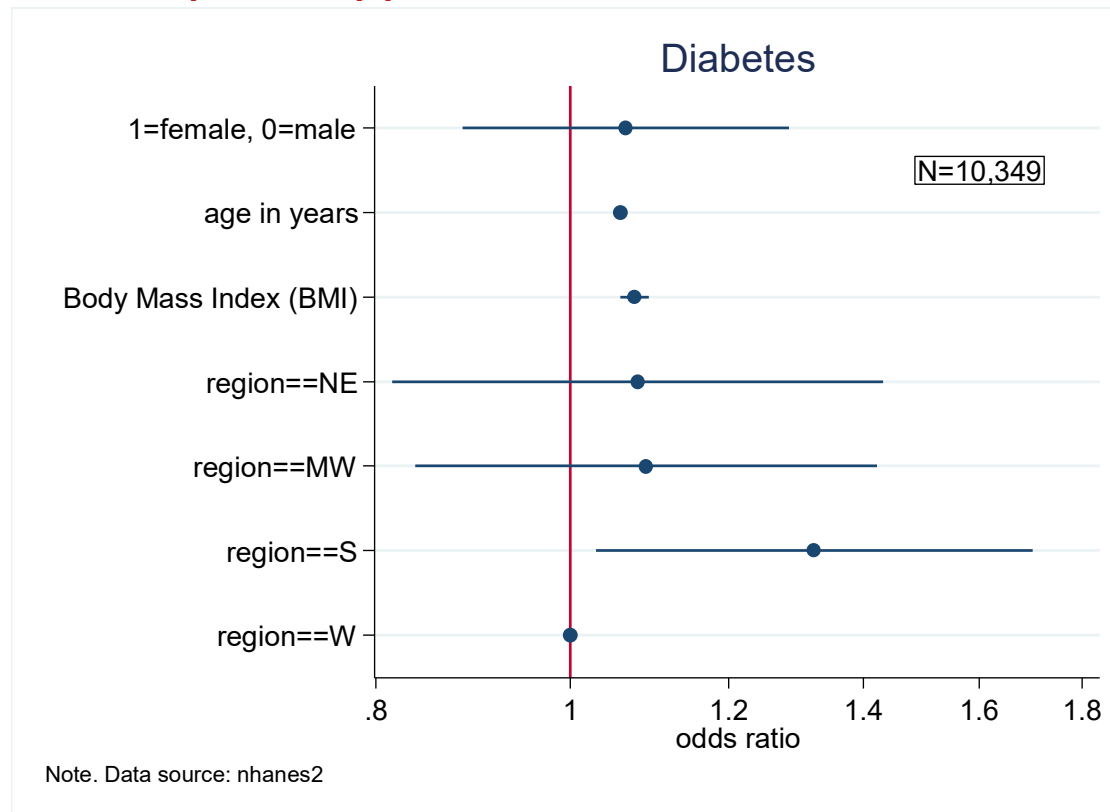


add titles
and text?

Basic Usage: Plotting a Single Model

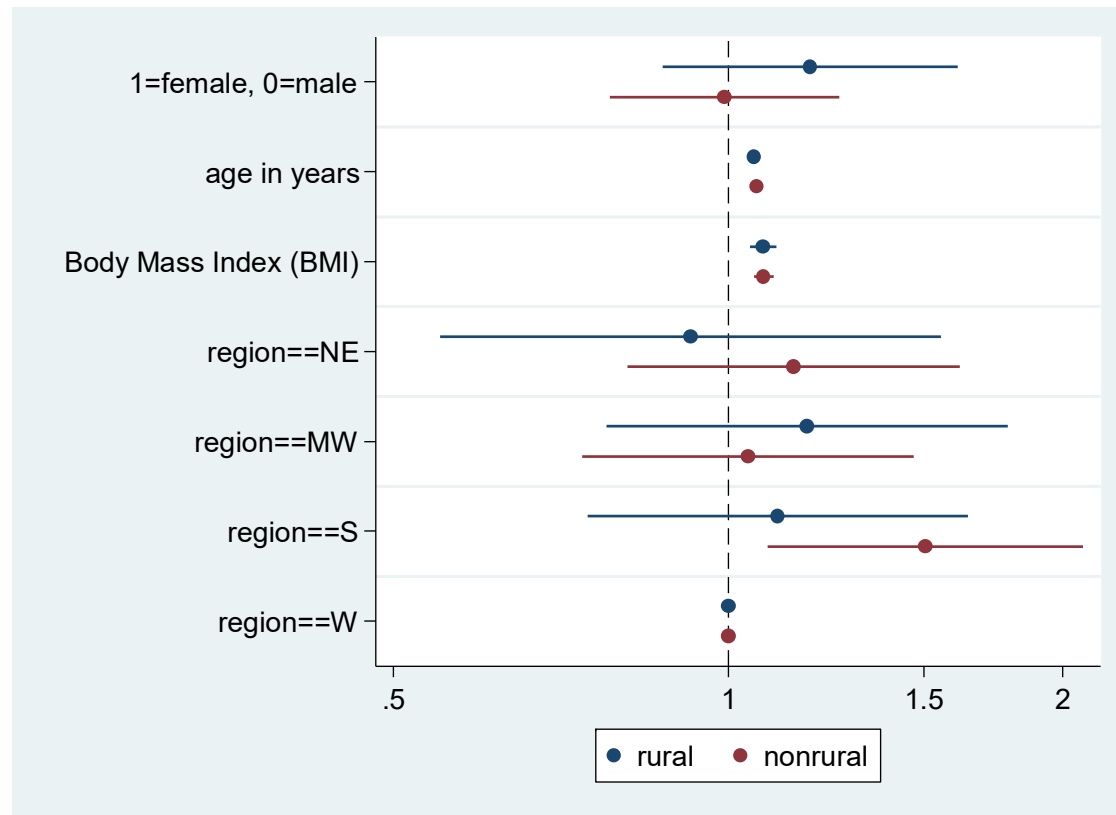
```
#delimit ;  
coefplot fem_age_bmi_reg,  
eform drop(_cons) xscale(log) xline(1) omitted  
title("Diabetes") xtitle("odds ratio")  
text(1.5 1.6 "N=10,349", box fcolor(white) lcolor(black))  
note("Note. Data source: nhanes2", span)  
graphregion(fcolor(white))
```

;



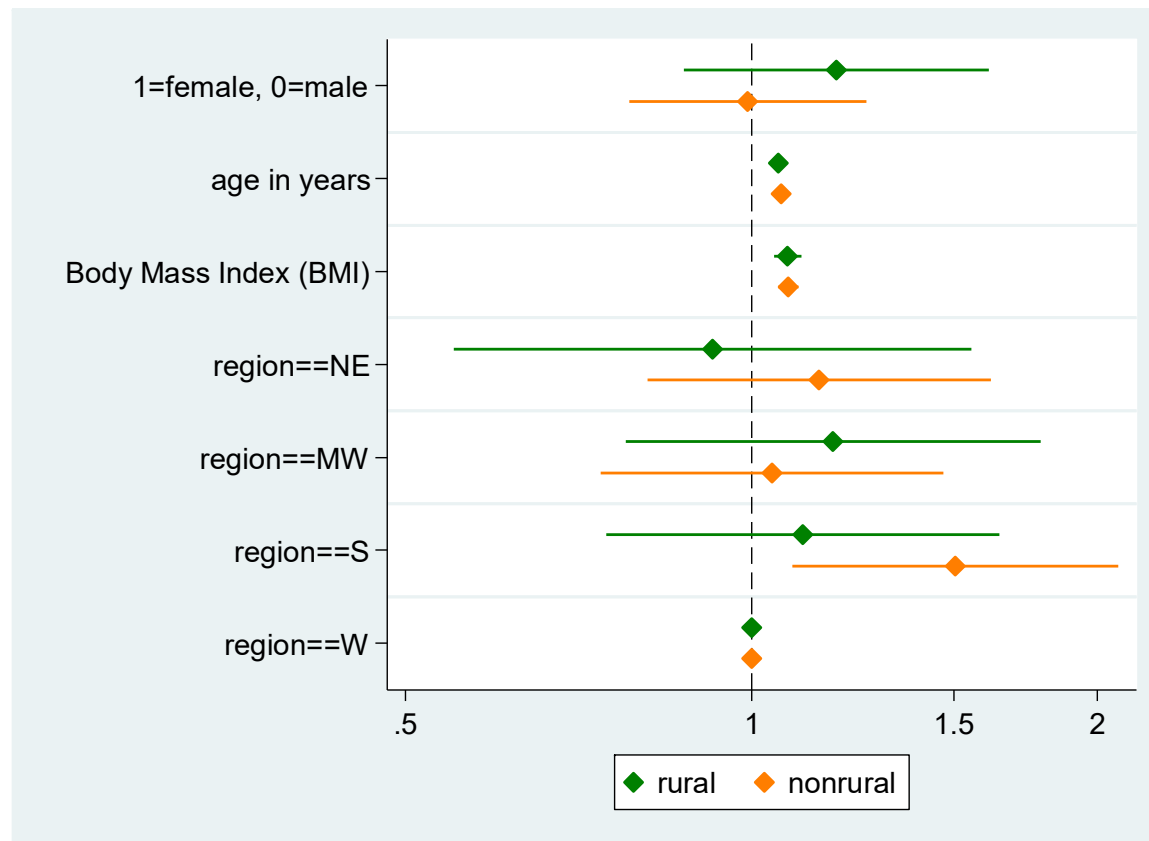
Basic Usage: Plotting Multiple Models

```
quietly logit diabetes female age bmi reg1 reg2 reg3 reg4 if rural == 1, or
estimates store rural
quietly logit diabetes female age bmi reg1 reg2 reg3 reg4 if rural == 0, or
estimates store nonrural
#delimit ;
coefplot rural nonrural,
eform drop(_cons) xscale(log range(.5 2)) omitted
xline(1, lcolor(black) lwidth(thin) lpattern(dash))
;
```



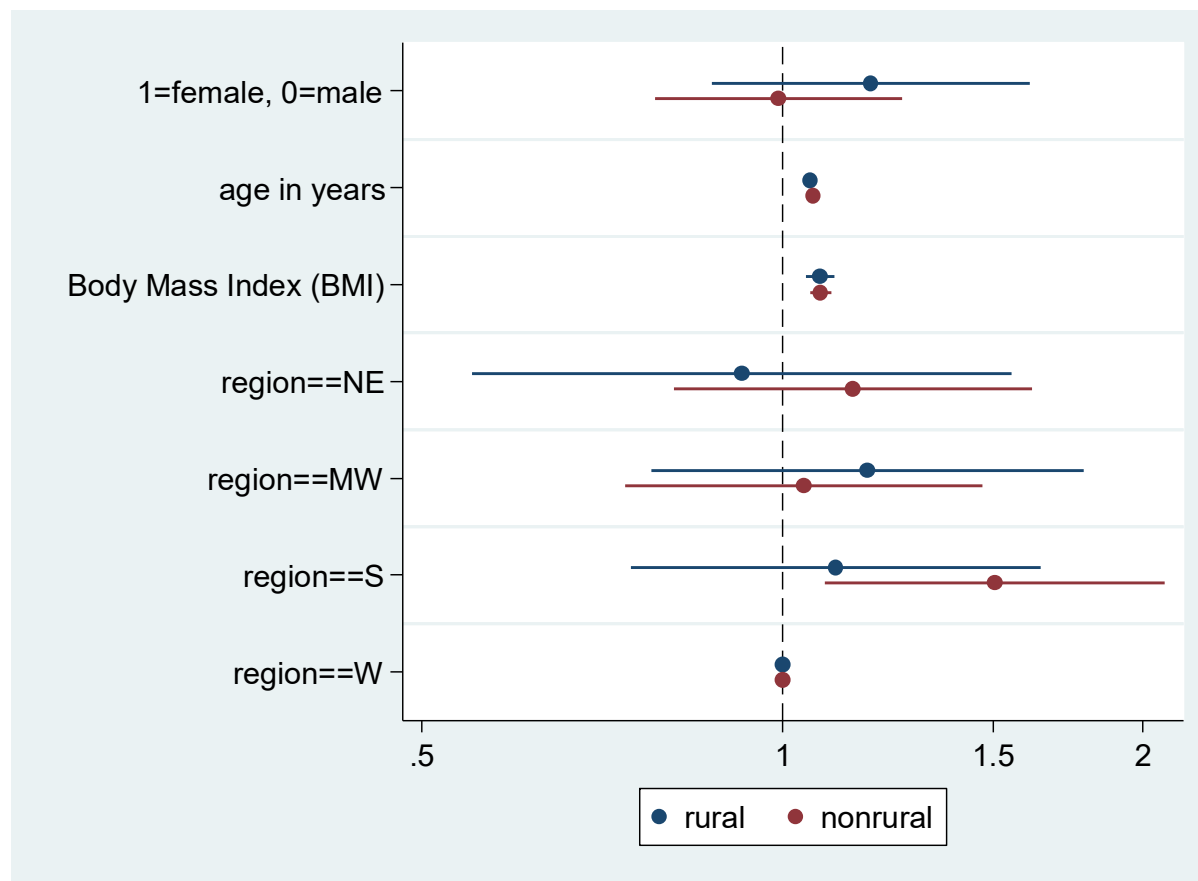
Basic Usage: Plotting Multiple Models

```
#delimit ;  
coefplot  
(rural, mcolor(green) ciopts(lcolor(green)))  
(nonrural, mcolor(orange) ciopts(lcolor(orange))),  
eform drop(_cons) xscale(log range(.5 2)) omitted  
xline(1, lcolor(black) lwidth(thin) lpattern(dash))  
msymbol(D)  
;
```



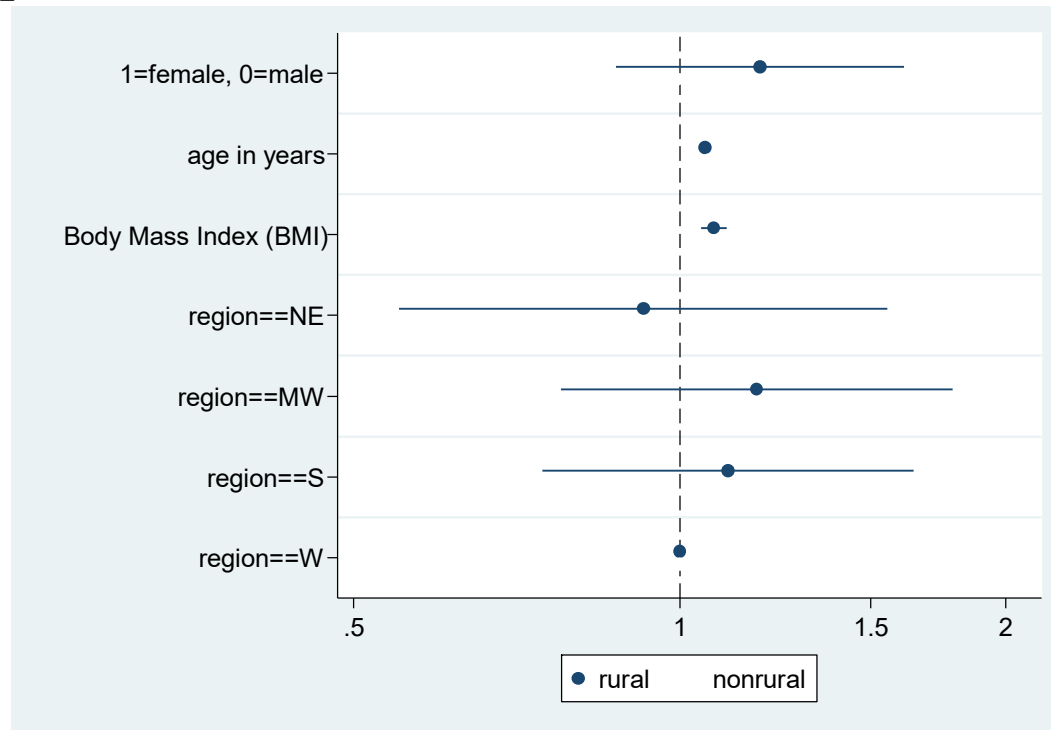
Basic Usage: Plotting Multiple Models

```
#delimit ;  
coefplot (rural, offset(.08)) (nonrural, offset(-.08)),  
eform drop(_cons) xscale(log range(.5 2)) omitted  
xline(1, lcolor(black) lwidth(thin) lpattern(dash))  
i
```



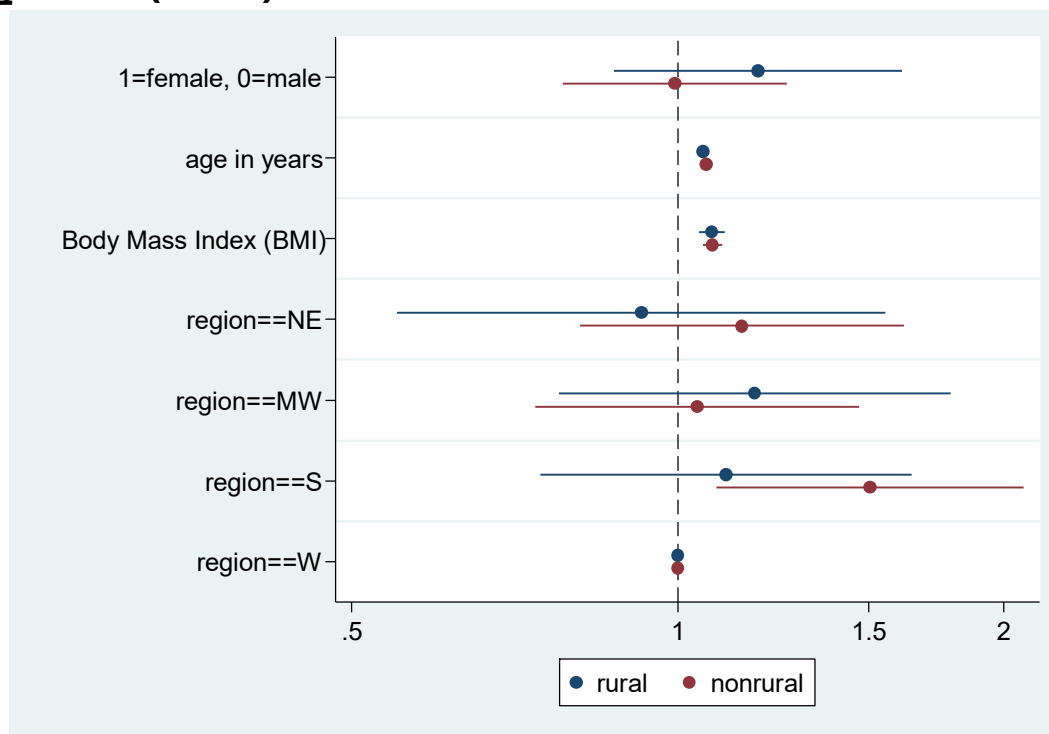
Basic Usage: Plotting Multiple Models

```
#delimit ;  
coefplot  
(rural, offset(.08))  
(nonrural, offset(-.08) mcolor(white) ciopts(lcolor(white))),  
eform drop(_cons) xscale(log range(.5 2)) omitted  
xline(1, lcolor(black) lwidth(thin) lpattern(dash))  
xsize(5.5) ysize(3.8)  
;
```



Basic Usage: Plotting Multiple Models

```
#delimit ;  
coefplot  
(rural, offset(.08))  
(nonrural, offset(-.08)),  
eform drop(_cons) xscale(log range(.5 2)) omitted  
xline(1, lcolor(black) lwidth(thin) lpattern(dash))  
xsize(5.5) ysize(3.8)  
;
```



Basic Usage: Subgraphs

```
#delimit ;  
quietly logit diabetes female age bmi reg1 reg2 reg3 reg4  
if rural == 1, or;  
estimates store rural;
```

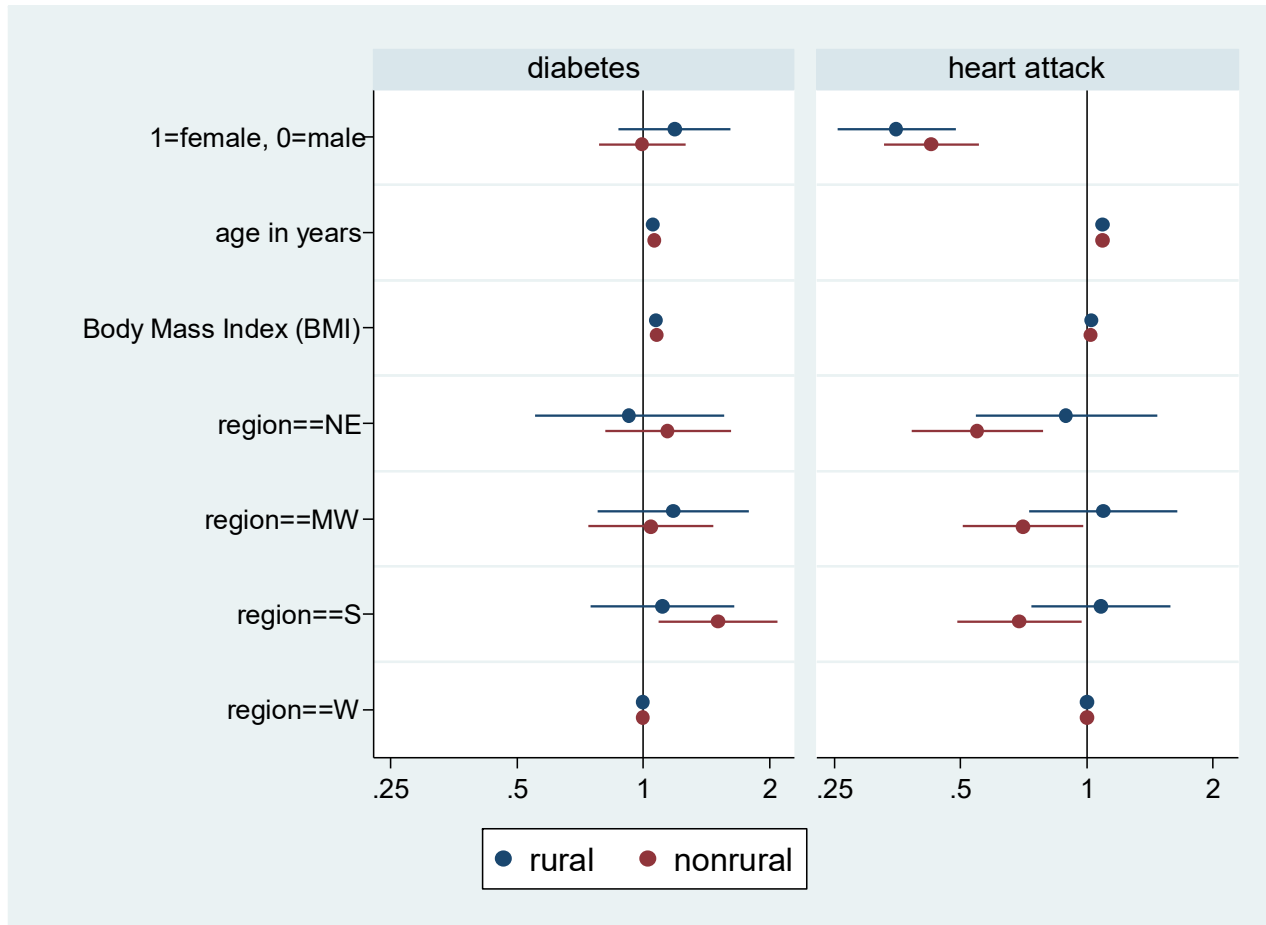
```
quietly logit diabetes female age bmi reg1 reg2 reg3 reg4  
if rural == 0, or;  
estimates store nonrural;
```

```
quietly logit heartatk female age bmi reg1 reg2 reg3 reg4  
if rural == 1, or;  
estimates store hrural;
```

```
quietly logit heartatk female age bmi reg1 reg2 reg3 reg4  
if rural == 0, or;  
estimates store hnonrural;  
;
```

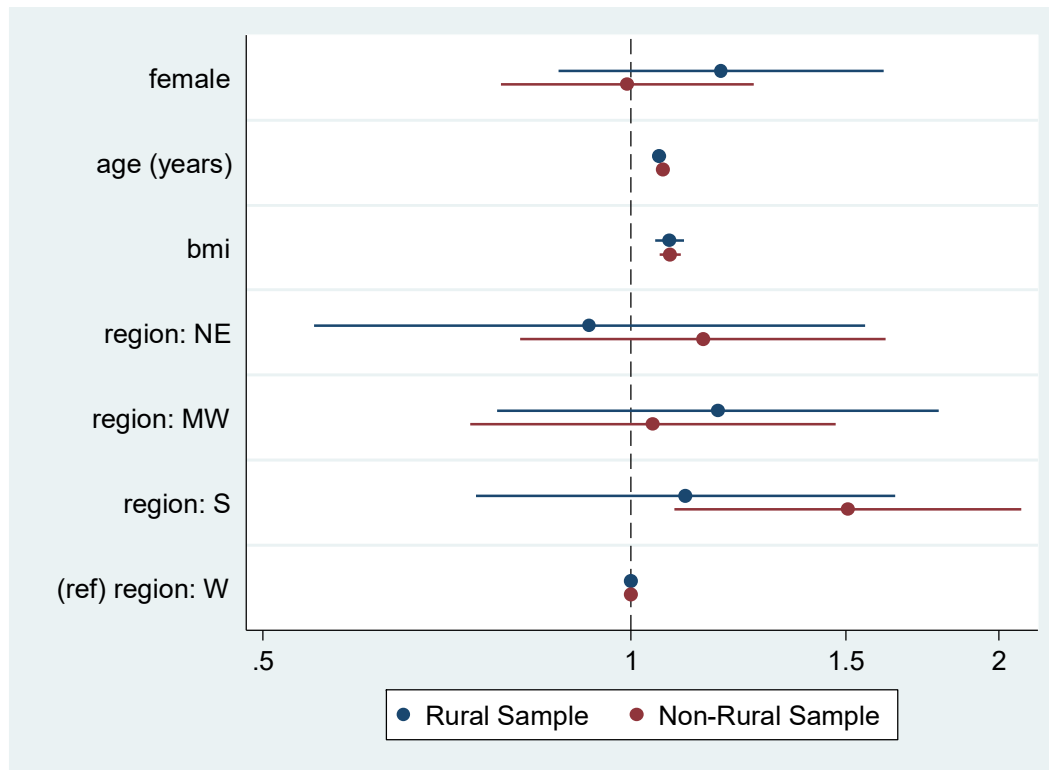
Basic Usage: Subgraphs

```
#delimit ;  
coefplot  
(rural, offset(.08)) (nonrural, offset(-.08)), bylabel(diabetes) ||  
(hrural, offset(.08)) (hnonrural, offset(-.08)), bylabel(heart attack) ||,  
eform drop(_cons) omitted xline(1, lcolor(black) lwidth(thin))  
xscale(log range(.25 2)) xlabel(.25 ".25" .5 ".5" 1 "1" 2 "2")  
;
```



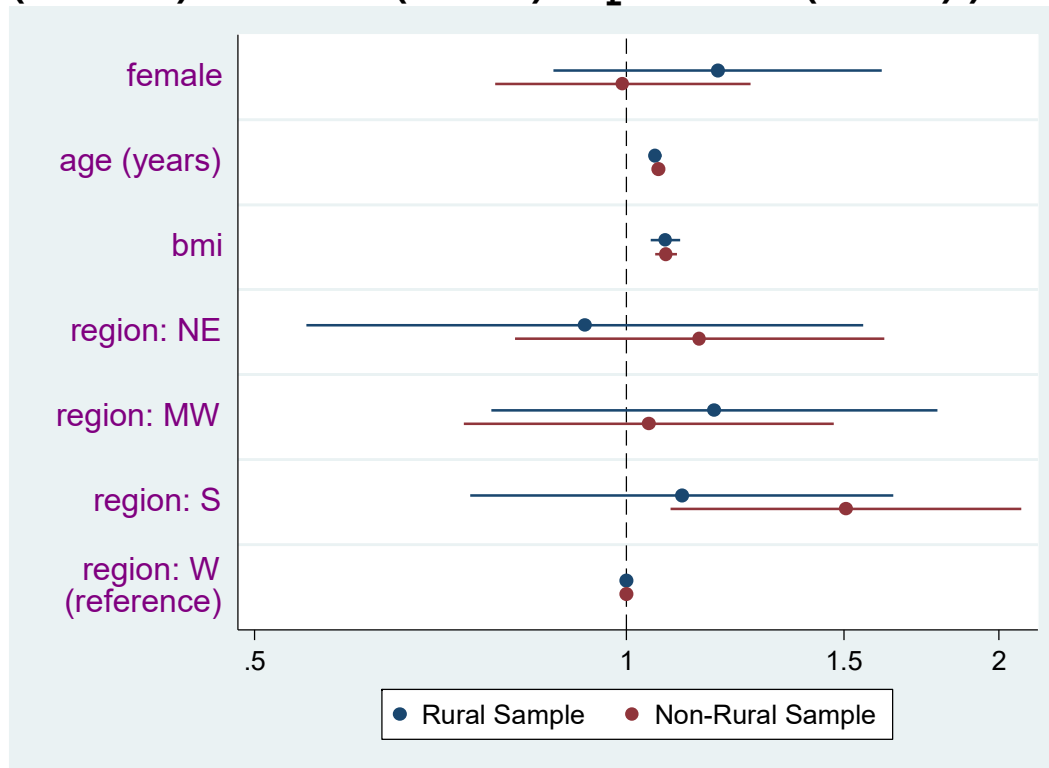
Labels

```
#delimit ;  
coefplot  
(rural, label(Rural Sample) offset(.08))  
(nonrural, label(Non-Rural Sample) offset(-.08)),  
coeflabels(female="female" age="age (years)" bmi="bmi"  
reg1="region: NE" reg2="region: MW" reg3="region: S"  
reg4="(ref) region: W", notick labgap(2))  
eform drop(_cons) xscale(log range(.5 2)) omitted  
xline(1, lcolor(black) lwidth(thin) lpattern(dash))  
;
```



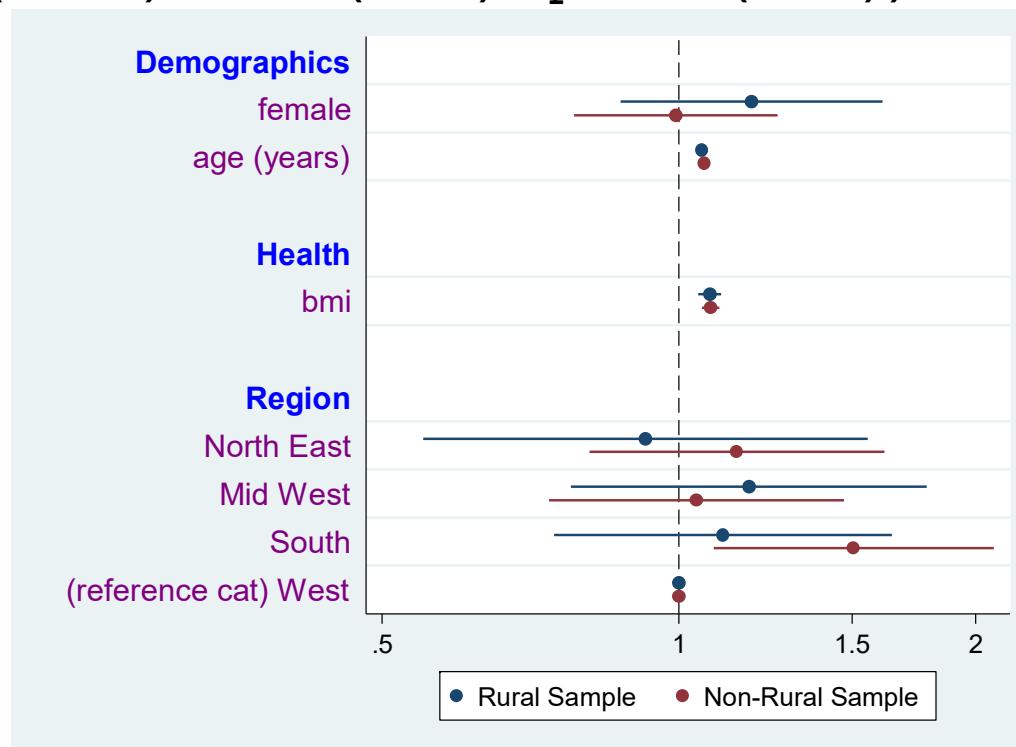
Labels

```
#delimit ;  
coefplot  
(rural, label(Rural Sample) offset(.08))  
(nonrural, label(Non-Rural Sample) offset(-.08)),  
coeflabels(female="female" age="age (years)" bmi="bmi"  
reg1="region: NE" reg2="region: MW" reg3="region: S"  
reg4="region: W (reference)",  
notick wrap(12) labszsize(medlarge) labcolor(purple) labgap(2))  
eform drop(_cons) xscale(log range(.5 2)) omitted  
xline(1, lcolor(black) lwidth(thin) lpattern(dash))  
;
```



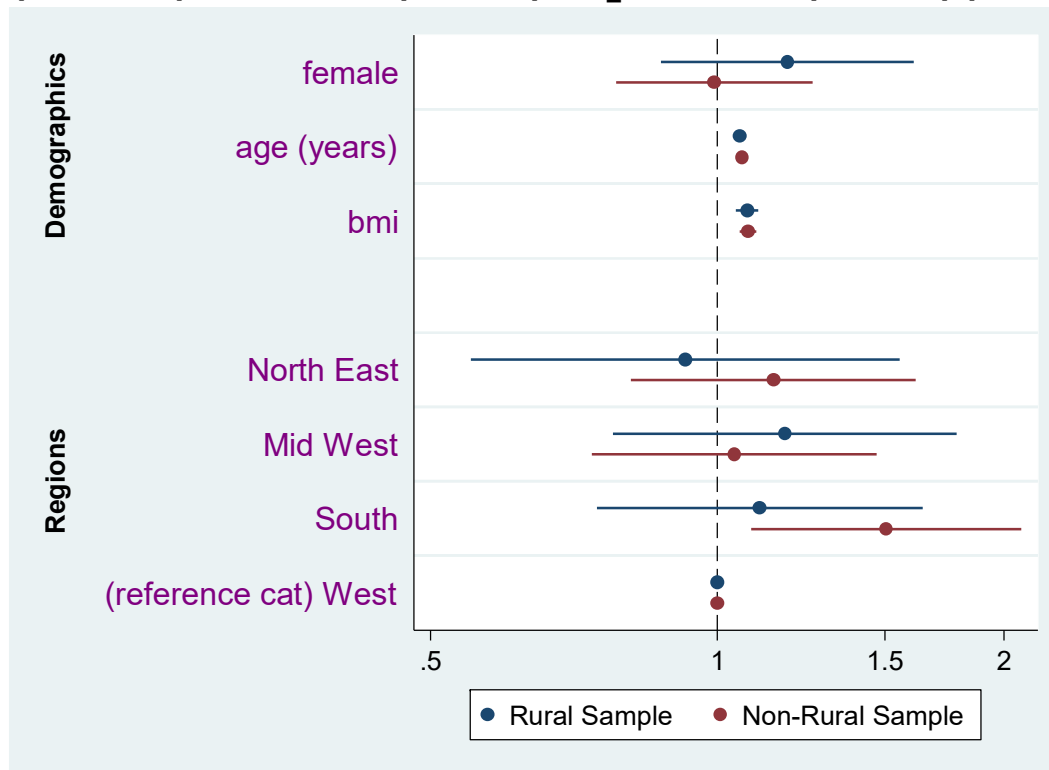
Labels

```
#delimit ;
coefplot
(rural, label(Rural Sample) offset(.14))
(nonrural, label(Non-Rural Sample) offset(-.14)),
coeflabels(female="female" age="age (years)" bmi="bmi"
reg1="North East" reg2="Mid West" reg3="South"
reg4="(reference cat) West",
notick labsize(medlarge) labcolor(purple) labgap(2))
headings(female="{bf:Demographics}" bmi="{bf:Health}"
reg1="{bf:Region}", labcolor(blue))
eform drop(_cons) xscale(log range(.5 2)) omitted
xline(1, lcolor(black) lwidth(thin) lpattern(dash))
;
```



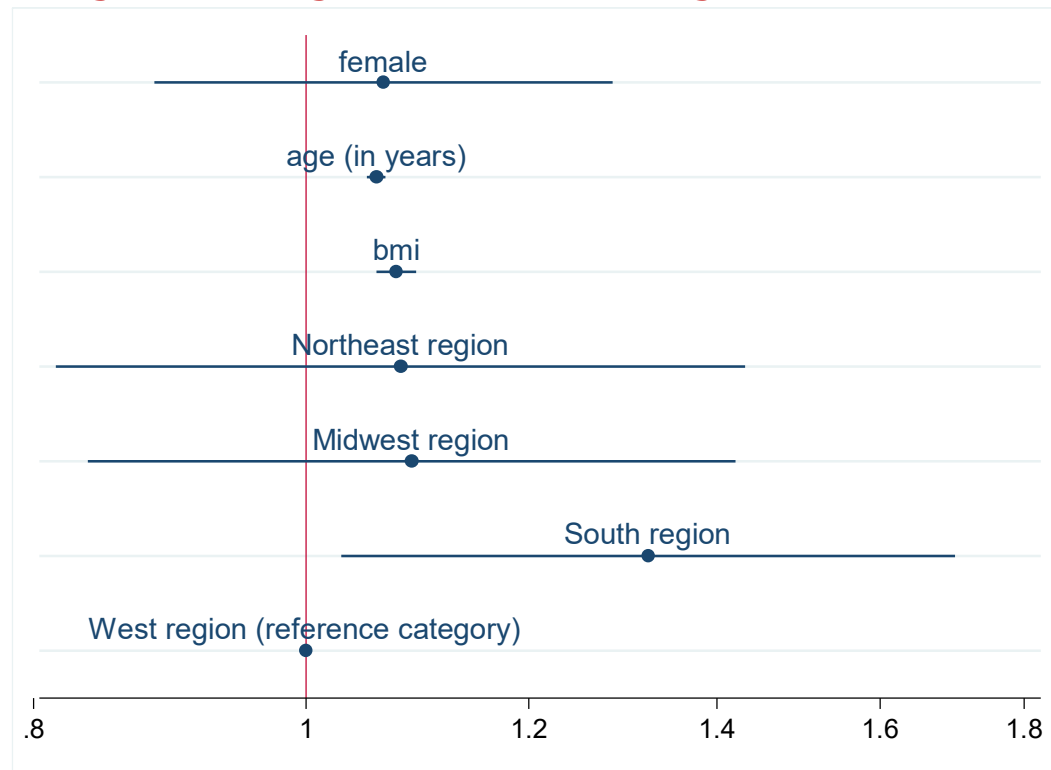
Labels

```
#delimit ;
coefplot
(rural, label(Rural Sample) offset(.14))
(nonrural, label(Non-Rural Sample) offset(-.14)),
coeflabels(female="female" age="age (years)" bmi="bmi"
reg1="North East" reg2="Mid West" reg3="South"
reg4="(reference cat) West",
notick labsize(medlarge) labcolor(purple) labgap(2))
groups(female age bmi="{bf:Demographics}" reg*="{bf:Regions}")
eform drop(_cons) xscale(log range(.5 2)) omitted
xline(1, lcolor(black) lwidth(thin) lpattern(dash))
;
```



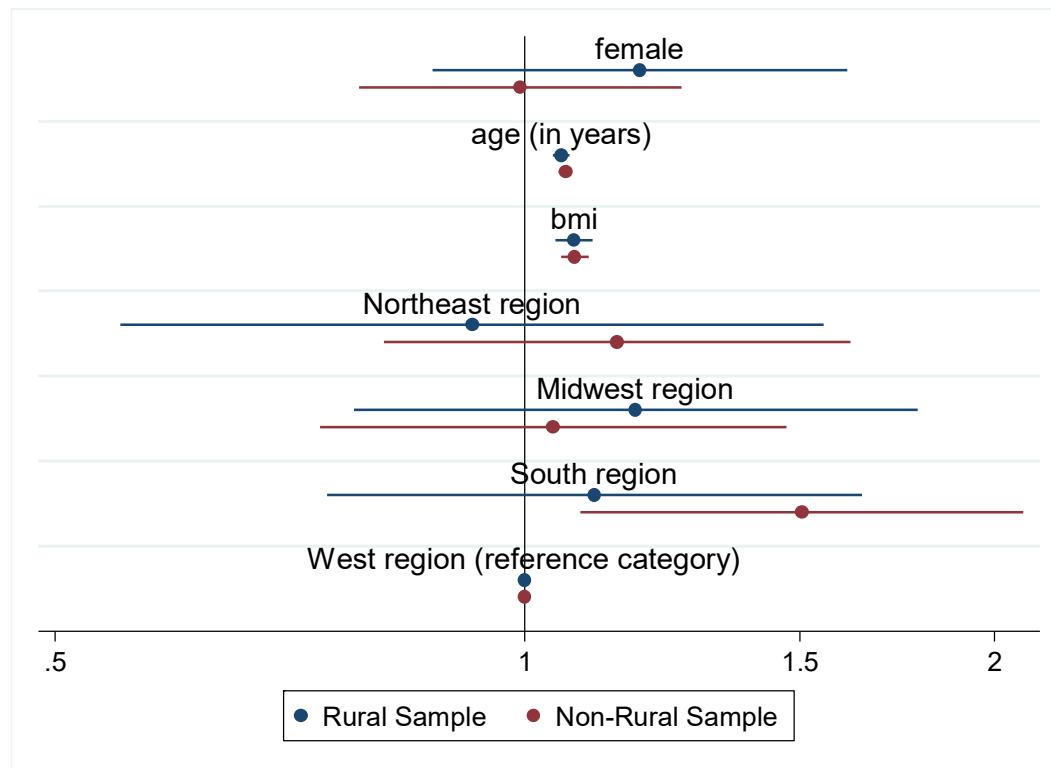
Labels: Markers

```
quietly logit diabetes female age bmi reg1 reg2 reg3 reg4, or
estimates store fem_age_bmi_reg
#delimit ;
coefplot fem_age_bmi_reg,
eform drop(_cons) xscale(log) xline(1, lwidth(vthin))
omitted yscale(off) mlabsize(medium) graphregion(fcolor(white))
mlabels(female=12 "female" age=12 "age (in years)" bmi=12 "bmi"
reg1=12 "Northeast region" reg2=12 "Midwest region"
reg3=12 "South region" reg4=12 "West region (reference category)")
i
```



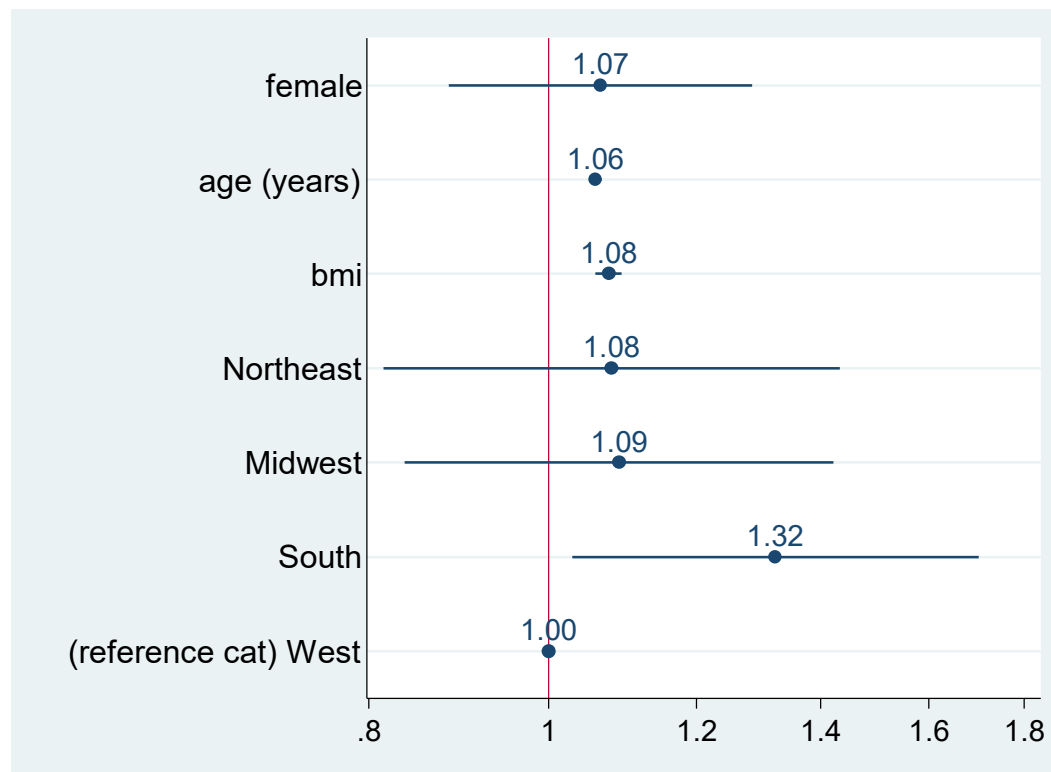
Labels: Markers

```
#delimit ;  
coefplot (rural, label(Rural Sample) offset(.1)  
mlabels(female=12 "female" age=12 "age (in years)" bmi=12 "bmi"  
reg1=12 "Northeast region" reg2=12 "Midwest region"  
reg3=12 "South region" reg4=12 "West region (reference category)")  
mlabsize(medium) mlabcolor(black))  
(nonrural, label(Non-Rural Sample) offset(-.1)),  
eform drop(_cons) xscale(log range(.5 2)) omitted yscale(off)  
xline(1, lwidth(vthin) lcolor(black)) graphregion(fcolor(white))  
;
```



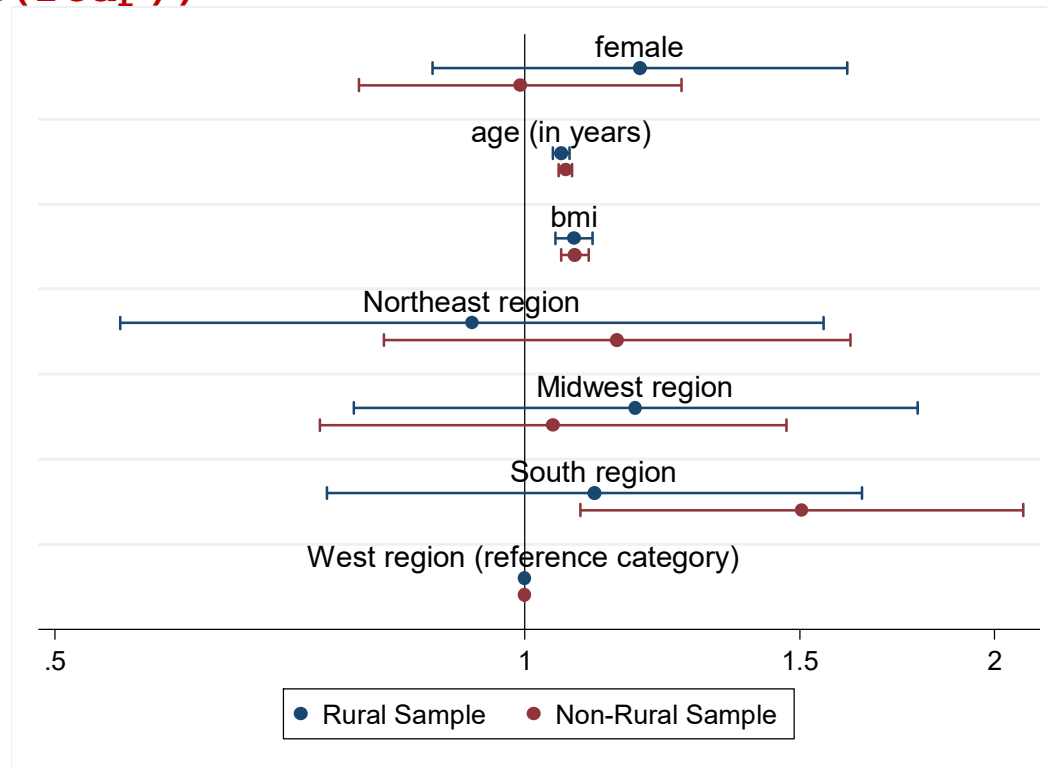
Labels: Markers

```
#delimit ;  
coefplot fem_age_bmi_reg,  
eform drop(_cons) xscale(log) xline(1, lwidth(vthin)) omitted  
mlabel format(%9.2f) mlabposition(12) mlabsize(medium)  
xlabel(.8 ".8" 1 "1" 1.2 "1.2" 1.4 "1.4" 1.6 "1.6" 1.8 "1.8",  
labsize(medium))  
coeflabels(female="female" age="age (years)" bmi="bmi"  
reg1="Northeast" reg2="Midwest" reg3="South"  
reg4="(reference cat) West", notick labsize(medlarge))  
;
```



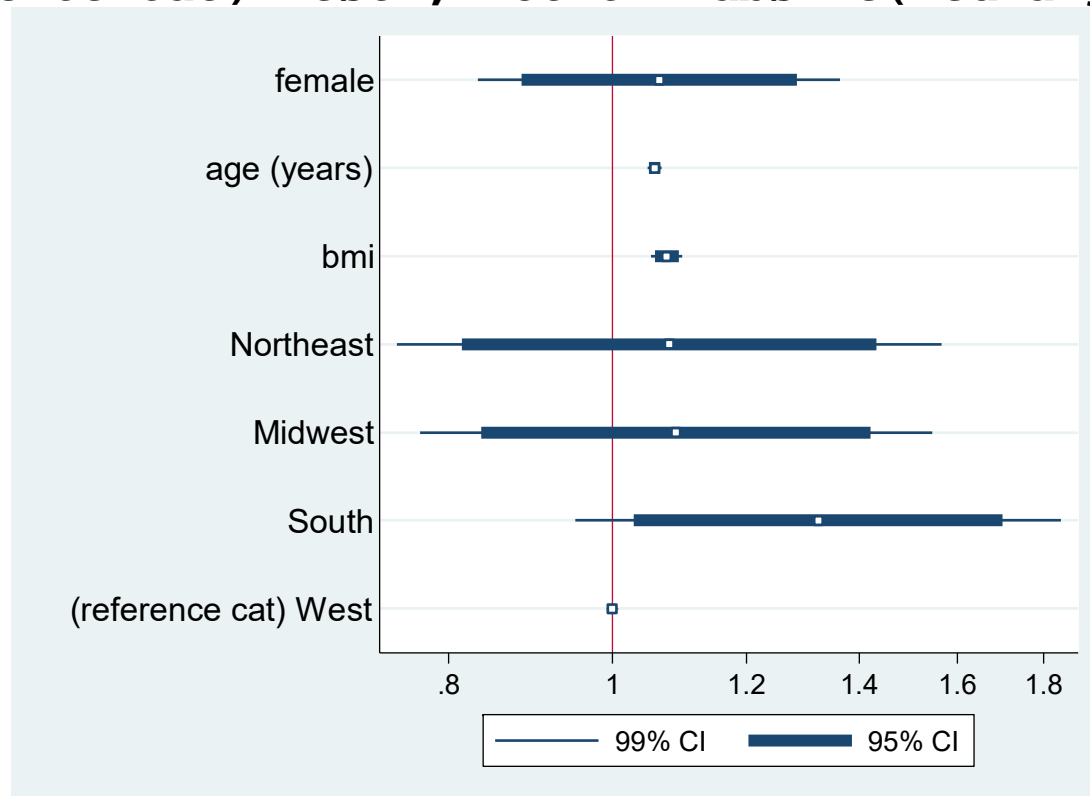
Confidence Intervals

```
#delimit ;  
coefplot (rural, label(Rural Sample) offset(.1)  
mlabels(female=12 "female" age=12 "age (in years)" bmi=12 "bmi"  
reg1=12 "Northeast region" reg2=12 "Midwest region"  
reg3=12 "South region" reg4=12 "West region (reference category)")  
mlabsize(medium) mlabcolor(black))  
(nonrural, label(Non-Rural Sample) offset(-.1)),  
eform drop(_cons) xscale(log range(.5 2)) omitted yscale(off)  
xline(1, lwidth(vthin) lcolor(black)) graphregion(fcolor(white))  
ciopts(recast(rcap))  
;
```



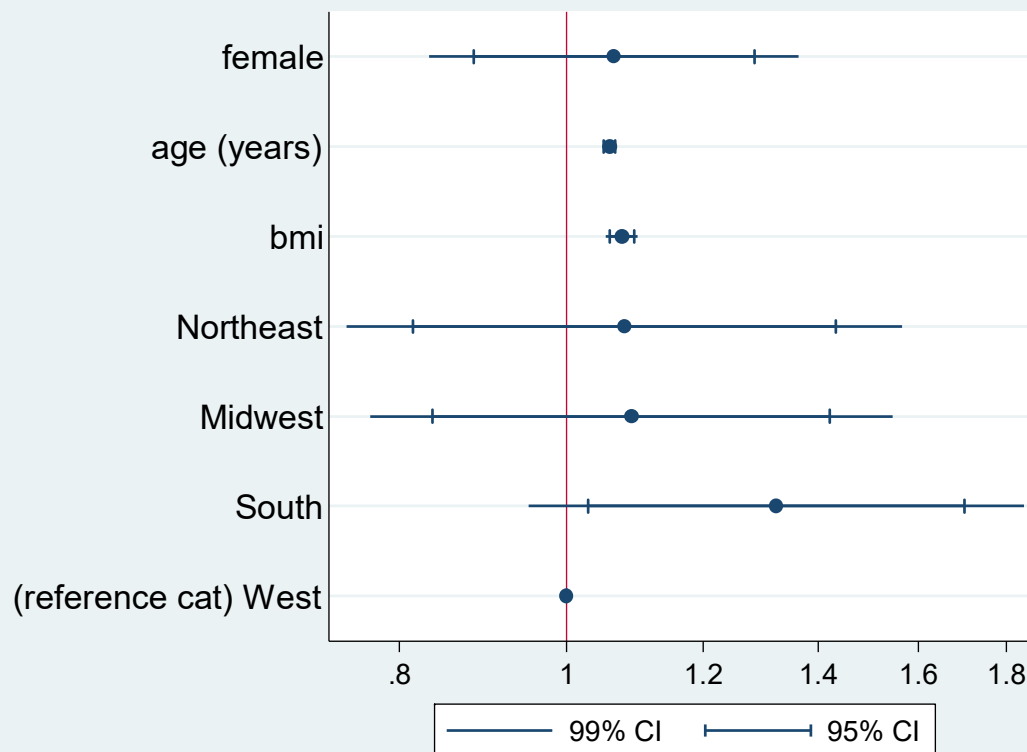
Confidence Intervals

```
#delimit ;  
coefplot fem_age_bmi_reg,  
eform drop(_cons) xscale(log) xline(1, lwidth(vthin)) omitted  
levels(99 95) msym(s) mfcolor(white)  
legend(order(1 "99% CI" 2 "95% CI")) ciopts(lwidth(*1 *5))  
coeflabels(female="female" age="age (years)" bmi="bmi"  
reg1="Northeast" reg2="Midwest" reg3="South"  
reg4="(reference cat) West", notick labszsize(medlarge))  
;
```



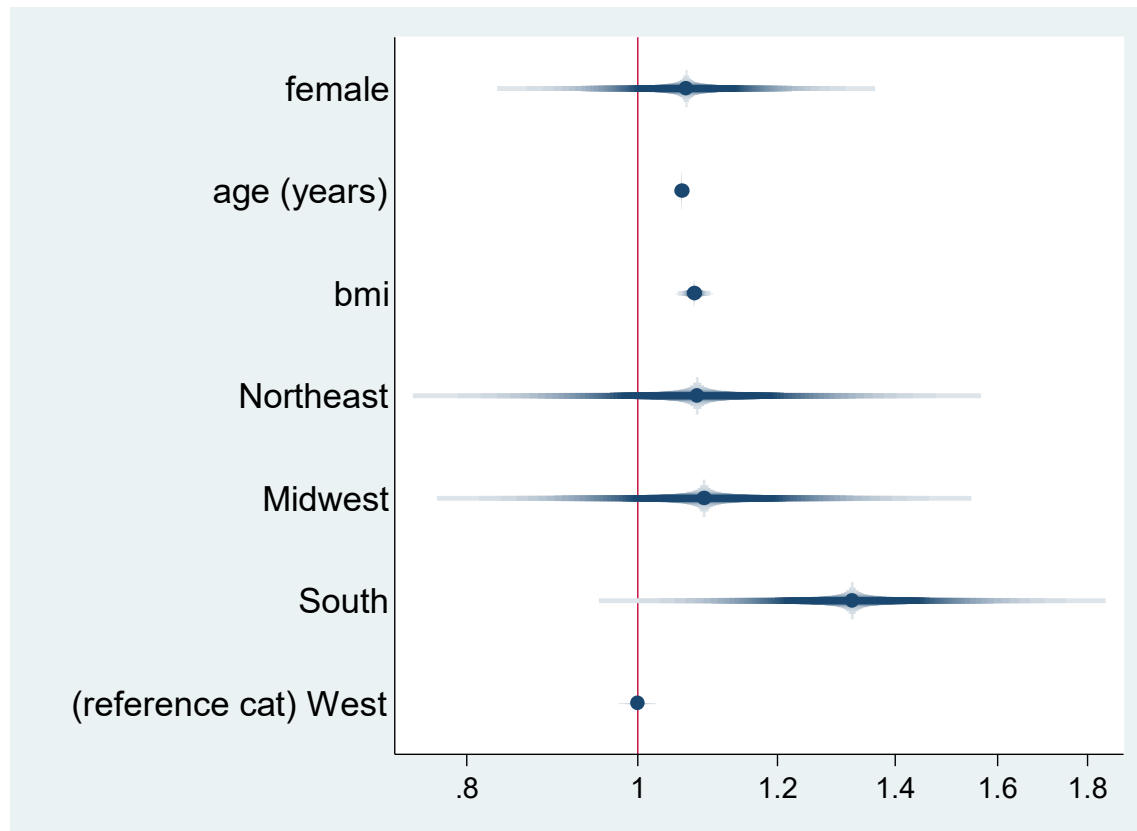
Confidence Intervals

```
#delimit ;  
coefplot fem_age_bmi_reg,  
eform drop(_cons) xscale(log) xline(1, lwidth(vthin)) omitted  
levels(99 95) legend(order(1 "99% CI" 2 "95% CI"))  
ciopts(recast(. rcap))  
coeflabels(female="female" age="age (years)" bmi="bmi"  
reg1="Northeast" reg2="Midwest" reg3="South"  
reg4="(reference cat) West", notick labszsize(medlarge))  
;
```



Confidence Intervals

```
#delimit ;  
coefplot fem_age_bmi_reg,  
eform drop(_cons) xscale(log) xline(1, lwidth(vthin)) omitted  
cismooth grid(none)  
coeflabels(female="female" age="age (years)" bmi="bmi"  
reg1="Northeast" reg2="Midwest" reg3="South"  
reg4="(reference cat) West", notick labszize(medlarge))  
;
```

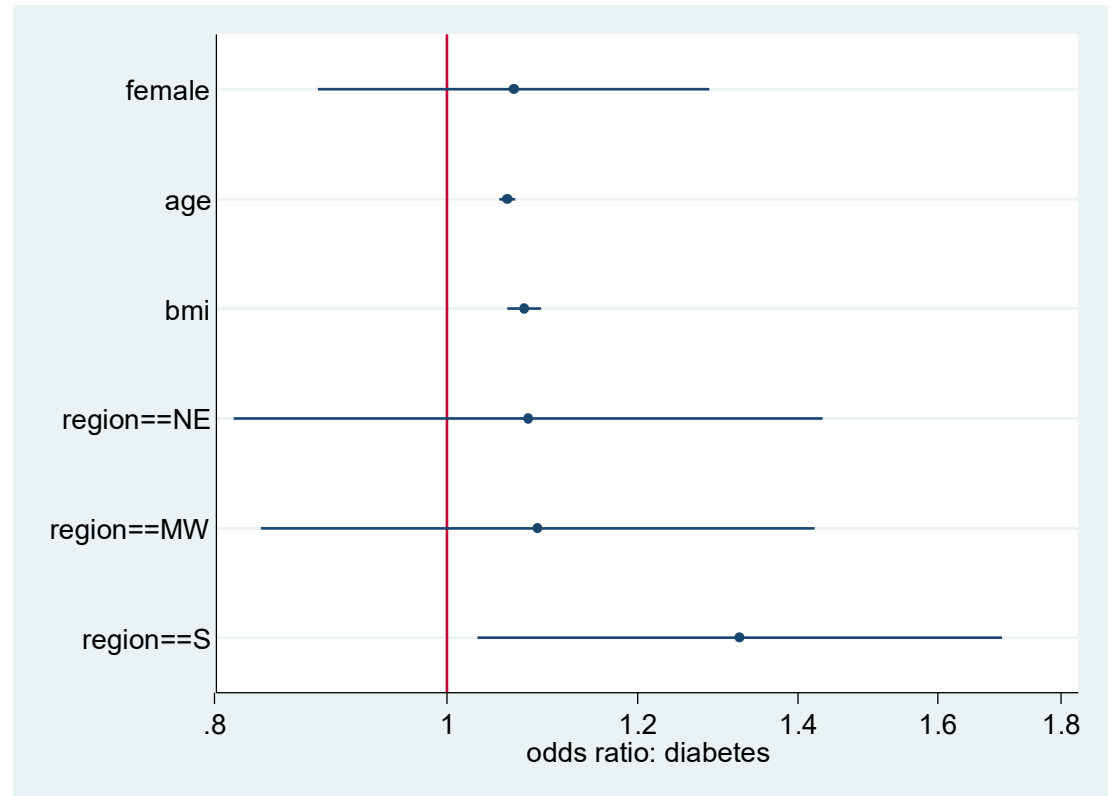


confidence intervals
for 50 equally spaced
levels (1, 3, ..., 99)
with graduated
color intensities and
varying line widths

The End ...
looking forward to seeing
more regression model results
displayed using **graphs!**

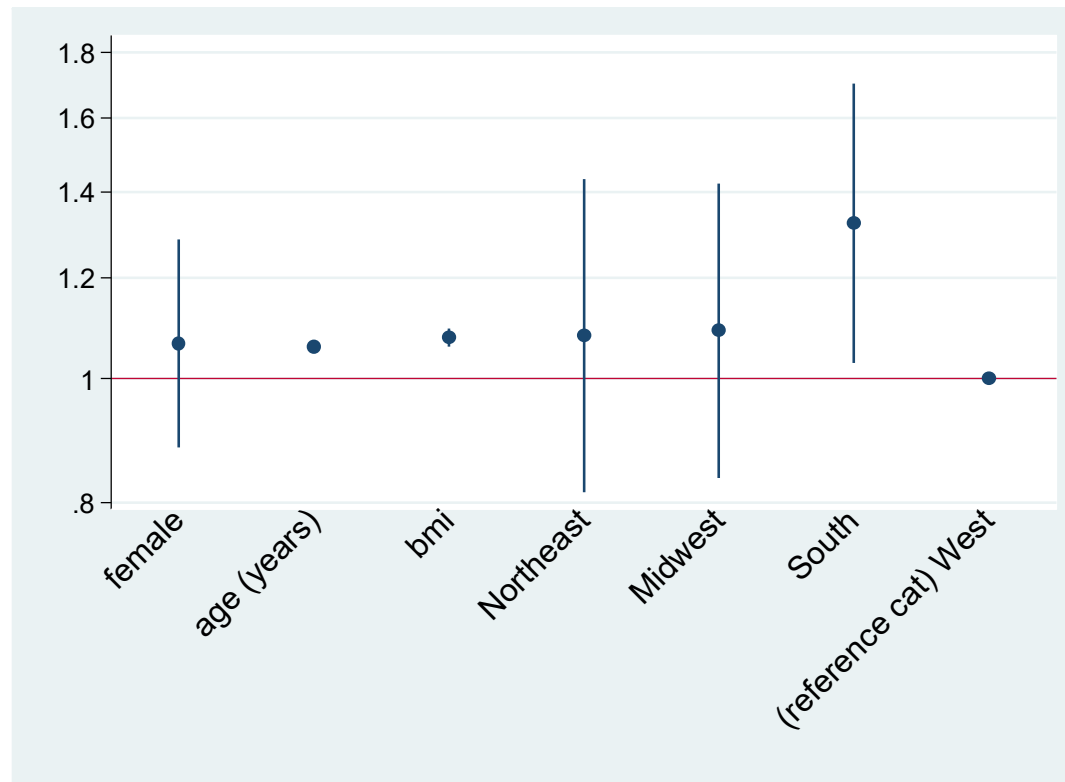
	diabetes
female	1.066 (0.102)
age	1.059*** (0.004)
bmi	1.077*** (0.009)
region==NE	1.081 (0.156)
region==MW	1.091 (0.148)
region==S	1.324** (0.170)

*** p<0.01, ** p<0.05, * p<0.1



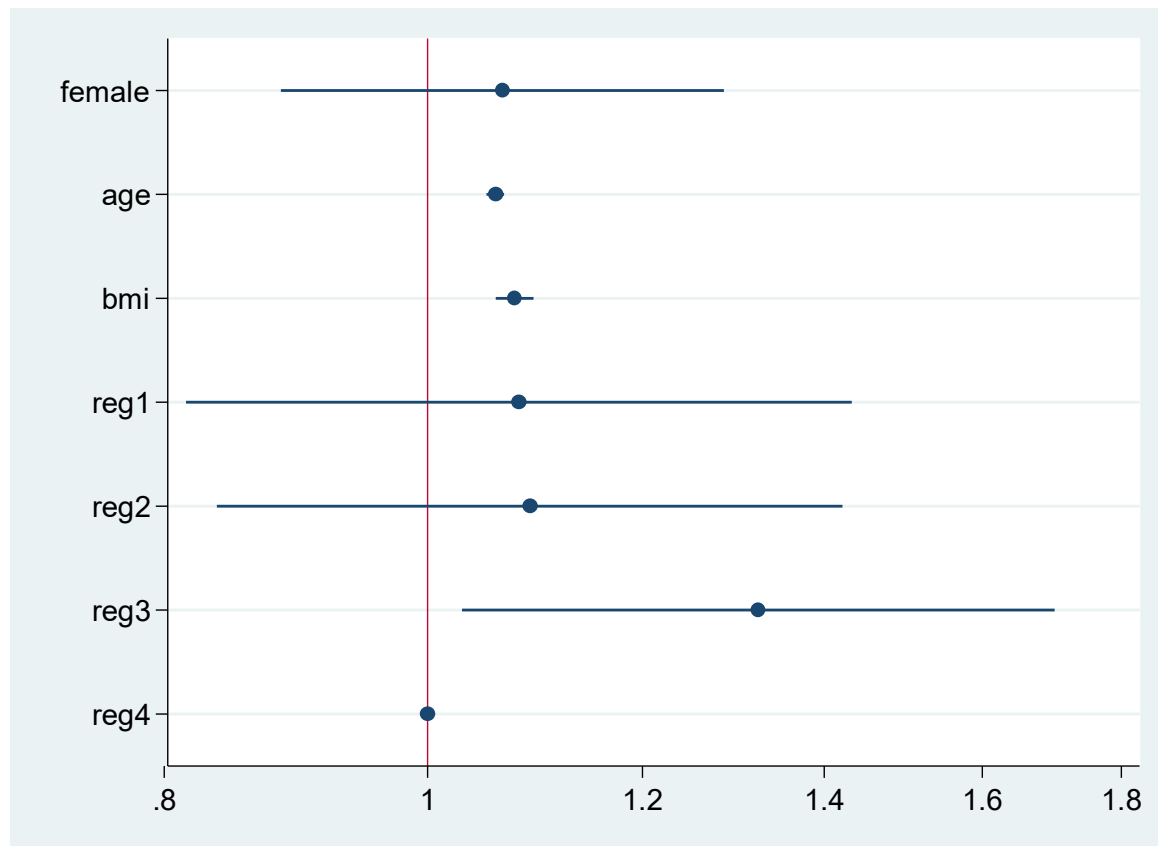
Additional

```
#delimit ;  
coefplot fem_age_bmi_reg,  
eform drop(_cons) omitted vertical  
yscale(log) yline(1, lwidth(vthin)) ylabel(,angle(0))  
xscale(lcolor(white)) xlabel(, angle(45))  
coeflabels(female="female" age="age (years)" bmi="bmi"  
reg1="Northeast" reg2="Midwest" reg3="South"  
reg4="(reference cat) West", notick labszsize(medlarge))  
;
```



Additional

```
#delimit ;  
coefplot fem_age_bmi_reg,  
eform drop(_cons) xscale(log) xline(1, lwidth(vthin)) omitted  
nolabels  
;
```



Additional

tab region, mi

1=NE, 2=MW, 3=S, 4=W	Freq.	Percent	Cum.
NE	2,096	20.25	20.25
MW	2,774	26.80	47.05
S	2,853	27.56	74.61
W	2,628	25.39	100.00
Total	10,351	100.00	

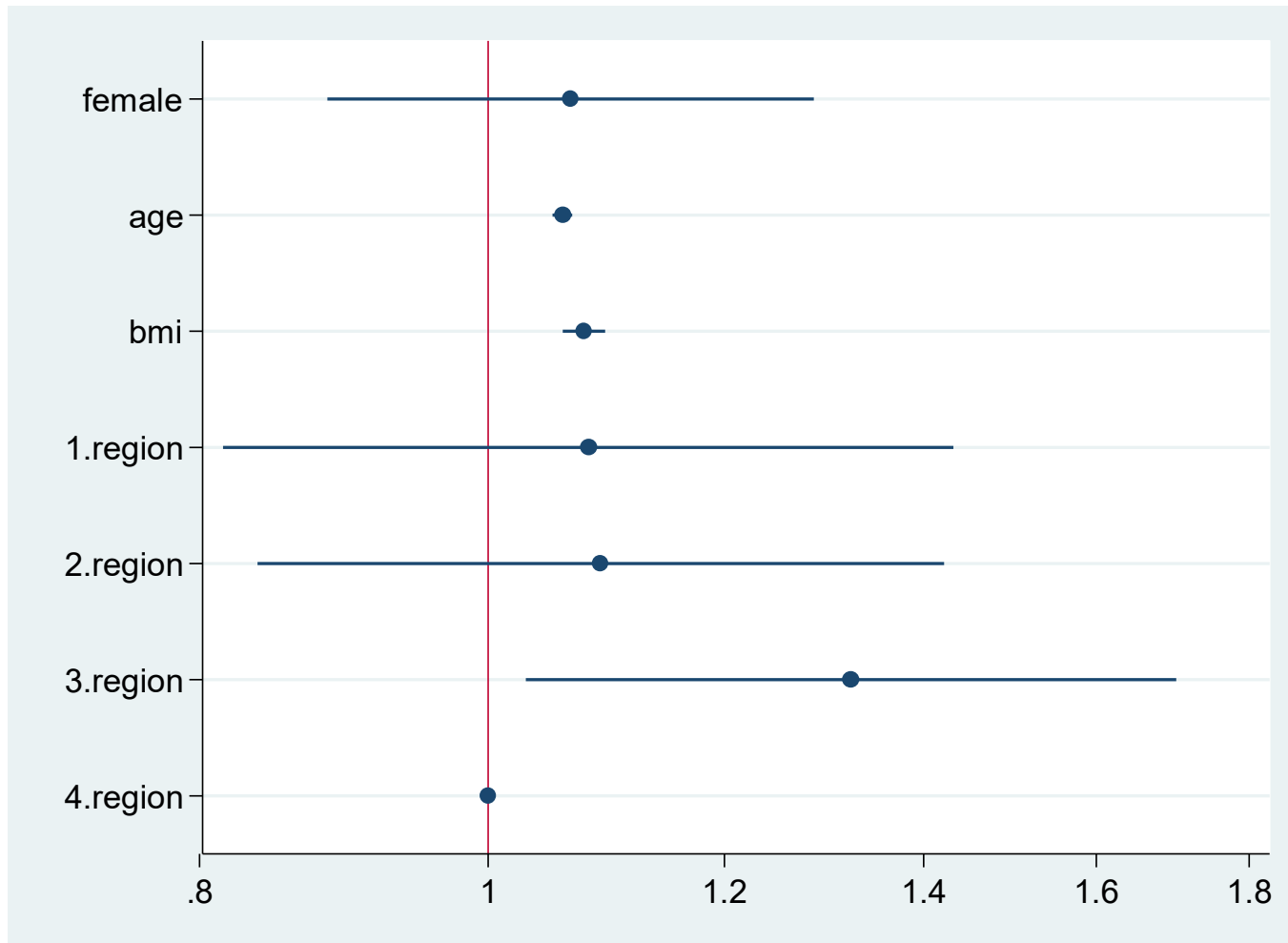
logit diabetes female age bmi **b4.region**, or

diabetes	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
female	1.0658	.1020453	0.67	0.506	.8834407 1.285803
age	1.059328	.0040456	15.09	0.000	1.051428 1.067287
bmi	1.076815	.0088939	8.96	0.000	1.059524 1.094389
region					
NE	1.080705	.1555003	0.54	0.590	.8151347 1.432797
MW	1.090745	.1475434	0.64	0.521	.8367247 1.421883
S	1.323504	.1696527	2.19	0.029	1.029472 1.701516

estimates store fem_age_bmi_reg_i

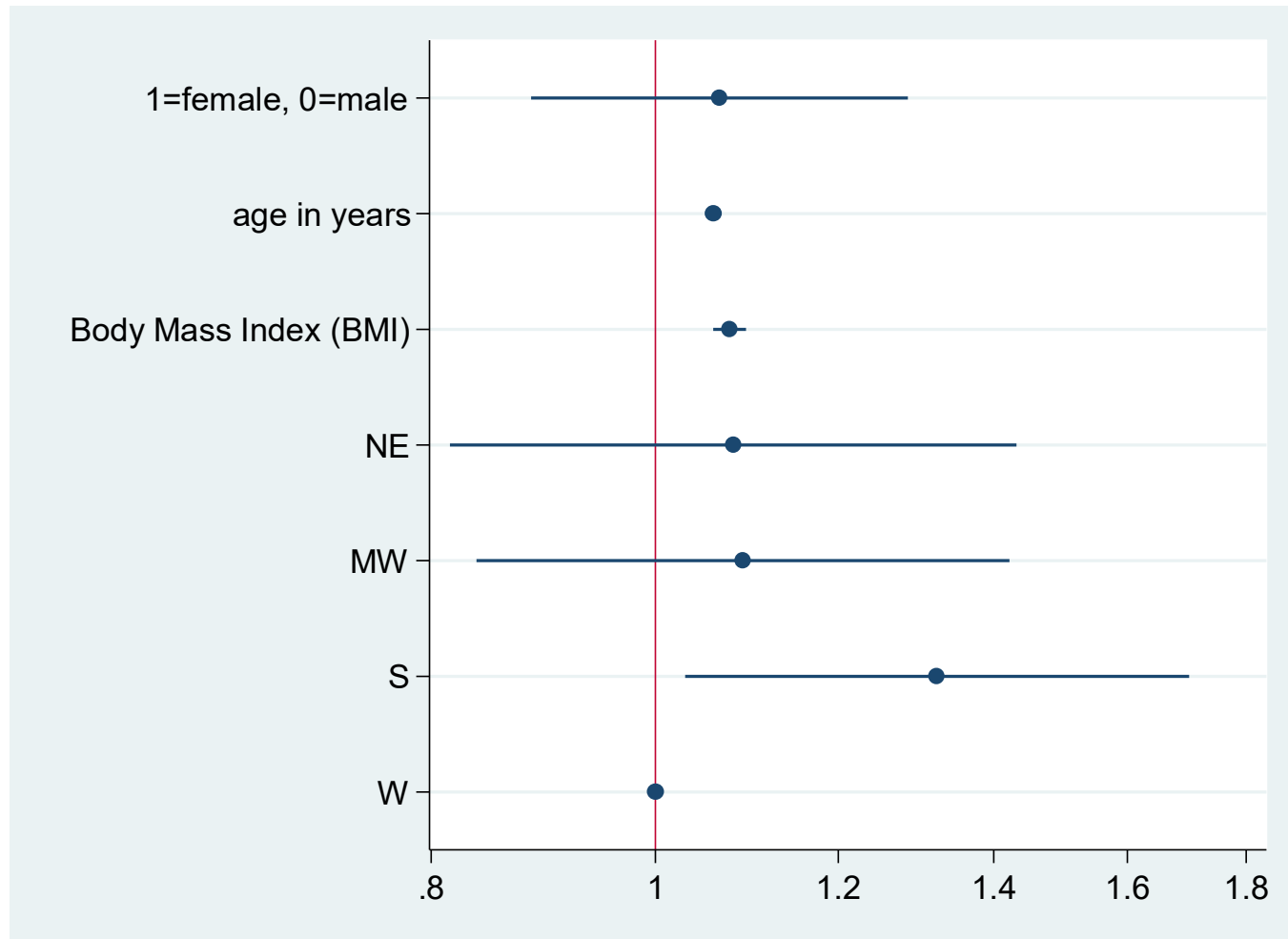
Additional

```
#delimit ;  
coefplot fem_age_bmi_reg_i,  
eform drop(_cons) xscale(log) xline(1, lwidth(vthin)) base  
nolabels  
i
```



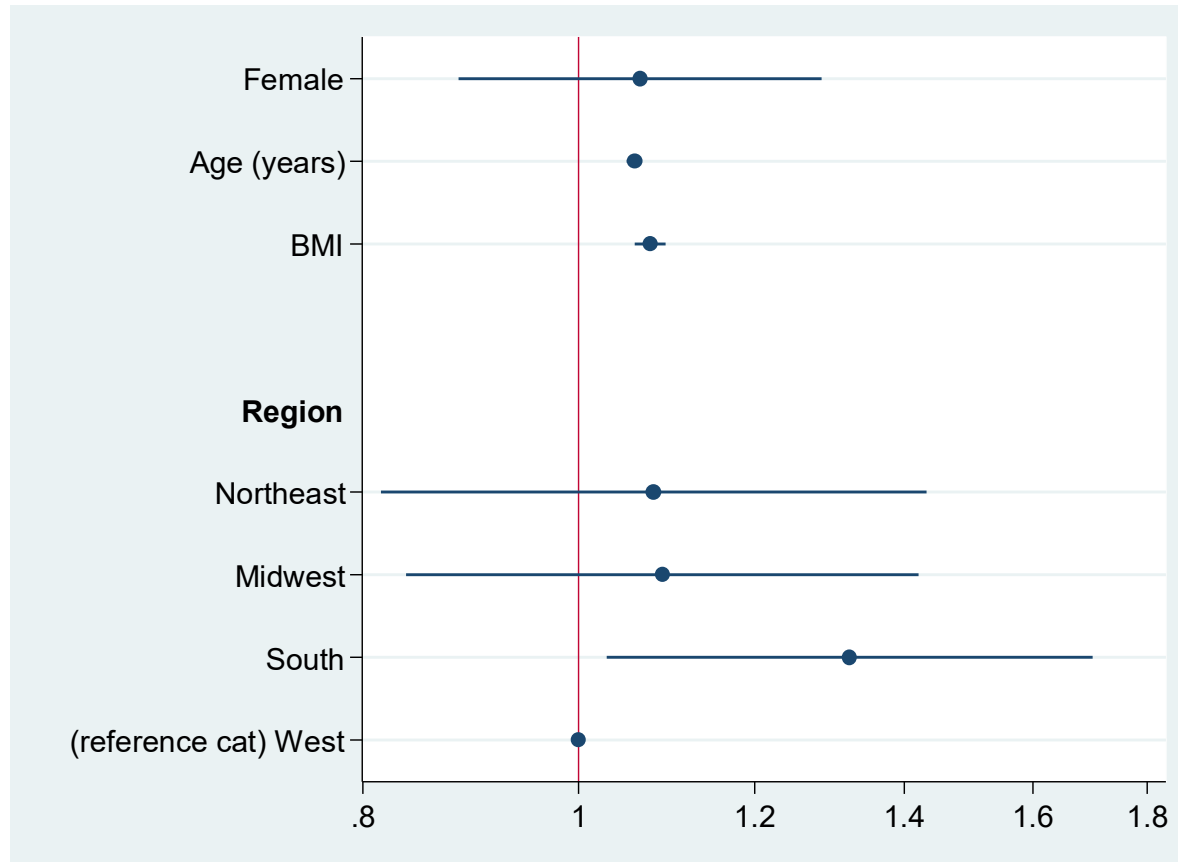
Additional

```
#delimit ;  
coefplot fem_age_bmi_reg_i,  
eform drop(_cons) xscale(log) xline(1, lwidth(vthin)) base  
;
```



Additional

```
#delimit ;  
coefplot fem_age_bmi_reg_i,  
eform drop(_cons) xscale(log) xline(1, lwidth(vthin)) base  
coeflabels(female="Female" age="Age (years)" bmi="BMI"  
1.region="Northeast" 2.region="Midwest" 3.region="South"  
4.region="(reference cat) West")  
headings(1.region="{bf:Region}")  
;
```



Additional

```
#delimit ;  
coefplot fem_age_bmi_reg_i,  
eform drop(_cons) xscale(log) xline(1, lwidth(vthin)) base  
coeflabels(female="Female" age="Age (years)" bmi="BMI"  
1.region="Northeast" 2.region="Midwest" 3.region="South"  
4.region="(reference cat) West")  
headings(1.region="{bf:Region}")  
order(bmi age female 4.region 1.region 2.region 3.region)  
;
```

