VIII Convegno Italiano degli Utenti di Stata

Sequential Logit Models: Transition probabilities among non alcoholic fatty liver disease (NAFLD) stages in a random sample population-based study from Southern Italy

Alberto R. Osella[1] María del Pilar Díaz[2]

[1]Laboratorio di Epidemiologia e Biostatistica
IRCCS "Saverio De Bellis"
Castellana (Bari) - Italia

[2]Cát. de Bioestadística. Facultad de Ciencias Médicas - Universidad Nacional de Córdoba. Córdoba, República Argentina

November 17th, 2011

Outline

Outline

Outline

Outline

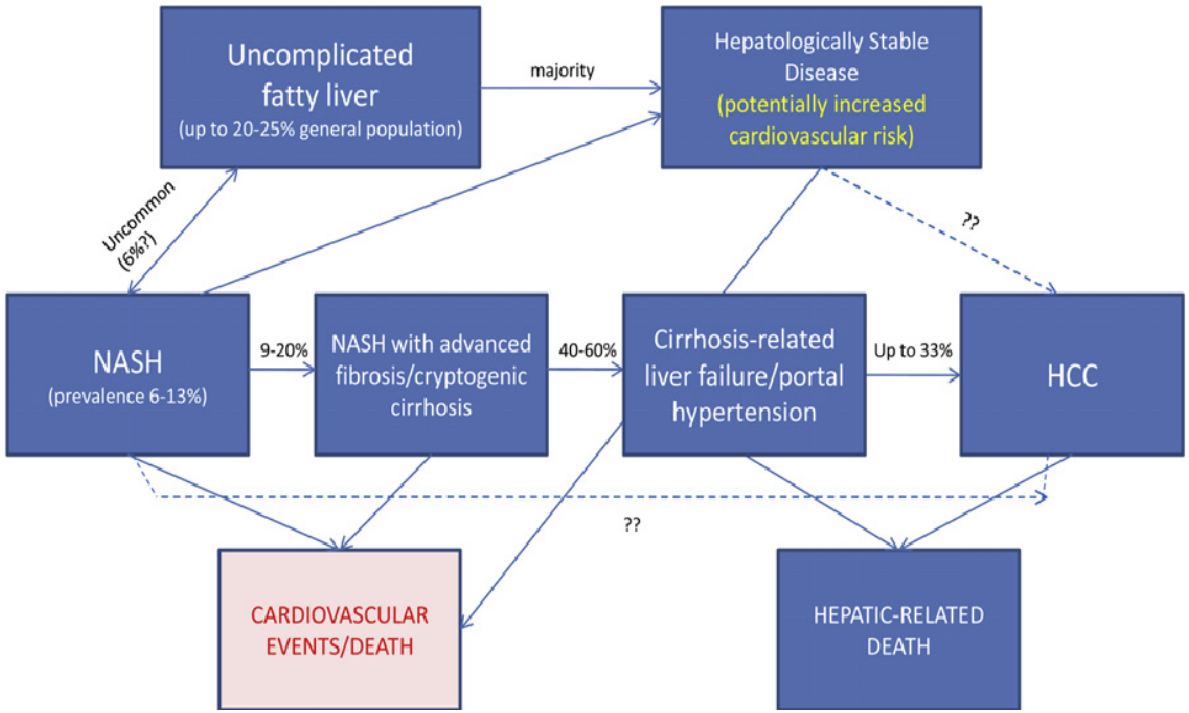
Outline

Definition

Fat accumulation in the liver in the absence of excessive alcohol consumption (less than 20 g per day) and any other specific causes of hepatic steatosis.

Source: Bacon BR et al. Nonalcoholic steatohepatitis: an expanded clinical entity. Gastroenterology 1994;107:1103-9

Natural History of NAFLD



Prevalence

NAFLD is now the most common hepatic disease worldwide. Its prevalence is increasing in the general population together with obesity, type 2 diabetes and the metabolic syndrome.

The NutriEP Study

Aim

To estimate liver disease and other health conditions prevalence in southern Italy: Hepatitis B, Hepatitis C, Overweight/Obesity and NAFLD.

The NutriEP Study

Design

Study Population: Putignano (BA). Inhabitants: 30.000
Population random sample: 2500 subjects.
Response rate 91% (1033 men and 1268 women were enrolled.
Study period: January 2006 to December 2007.

The NutriEP Study

Prevalence

Overweight: 34.5%
Obesity: 16.1%
NAFLD was present in 43.8% and 39% of overweight and obese subjects respectively.

The NutriEP Study

- Which is the impact of BMI on NAFLD in this mediterranean geographical area?
- Is the impact of BMI equal in all stages of NAFLD?

Model Names

- Sequential logit model (Mare, 1981)
- Sequential response model (Maddala, 1983)
- Mare model (Shavit and Blossfeld, 1993)
- Model for nested dichotomies (Fox, 1997)
- Continuation ration logit (Agresti, 2002)

The statistical model

- -seqlogit- fits a sequential logit model.
- It tests hypothesis across transitions.
- It implements the decomposition of the effect of a variable on the highest level of the dependent variable.
- It implements a sensitivity analysis.

Sequential Model

- to estimate the effect of the explanatory variables on the odds and probabilities of passing a set of transitions,
- each transition is modeled as a logistic regression using the sample which is 'at risk',

$$\widehat{p_{1i}} = \frac{exp(\alpha_1 + \lambda_1 BMI_i + \beta_1 x_i)}{1 + exp(\alpha_1 + \lambda_1 BMI_i + \beta_1 x_i)}$$

$$\widehat{p_{2i}} = \frac{exp(\alpha_2 + \lambda_2 BMI_i + \beta_2 x_i)}{1 + exp(\alpha_2 + \lambda_2 BMI_i + \beta_2 x_i)}$$

if $passing_{1i} = 1$

$$\widehat{p_{3i}} = \frac{exp(\alpha_3 + \lambda_3 BMI_i + \beta_3 x_i)}{1 + exp(\alpha_3 + \lambda_3 BMI_1 + \beta_3 x_i)}$$

if $passing_{2i} = 1$

Sequential Model

Maximun Expected Value of the Variable of Interest on the Outcome

$$E(L_i) = (1-\widehat{p_{1i}})l_0 + \widehat{p_{1i}}(1-p_{2i})l_1 + \widehat{p_{1i}}\widehat{p_{2i}}\widehat{(1-p_{3i})}l_2 + \widehat{p_{1i}}\widehat{p_{2i}}\widehat{p_3}i l_3$$

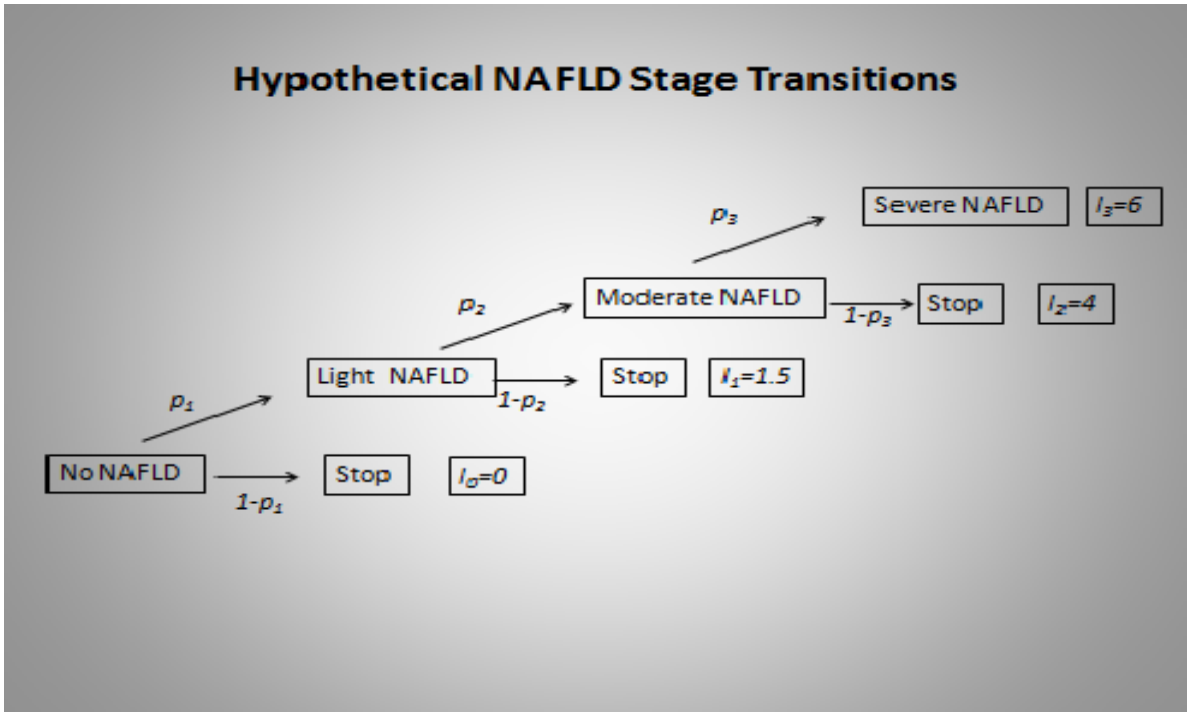
Testing assumption

- The exposure is not a prognostic factor:
  Mean duration of NAFLD is identical for exposed and unexposed subjects.
- The disease does not affect the exposure status.

The -seqlogit- command

```
seqlogit depvar [indepvars] [if] [in] [weight] ,
tree(tree)
[ofinterest(varname) over(varlist) sd(numlist)
deltasd(varname numlist) rho(#)
{ pr(numlist) | mn(# # , # # [, # #, etc.]) |
uniform } draws(#) drawstart(#) or
constraints(numlist) robust
cluster(clustervar) nolog level(#) maximize_options
by ... : may be used with seqlogit; see help by.
pweights, fweights and iweights are allowed;
see help weights.
```

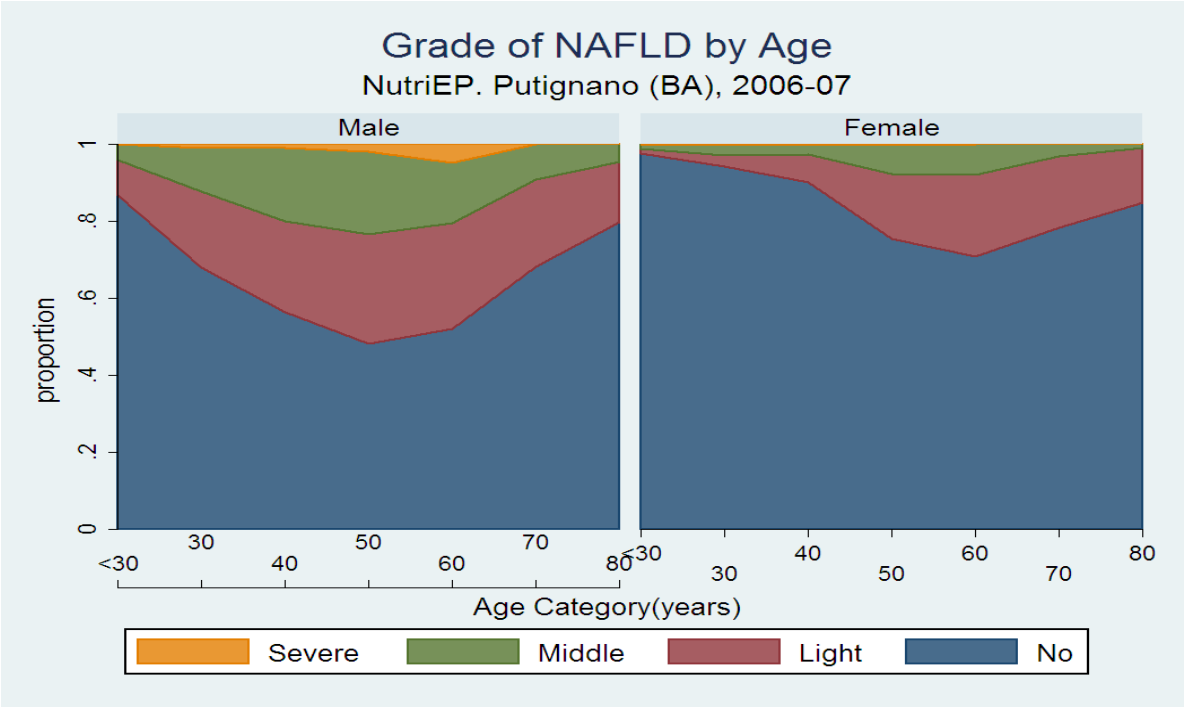
Graphical Model



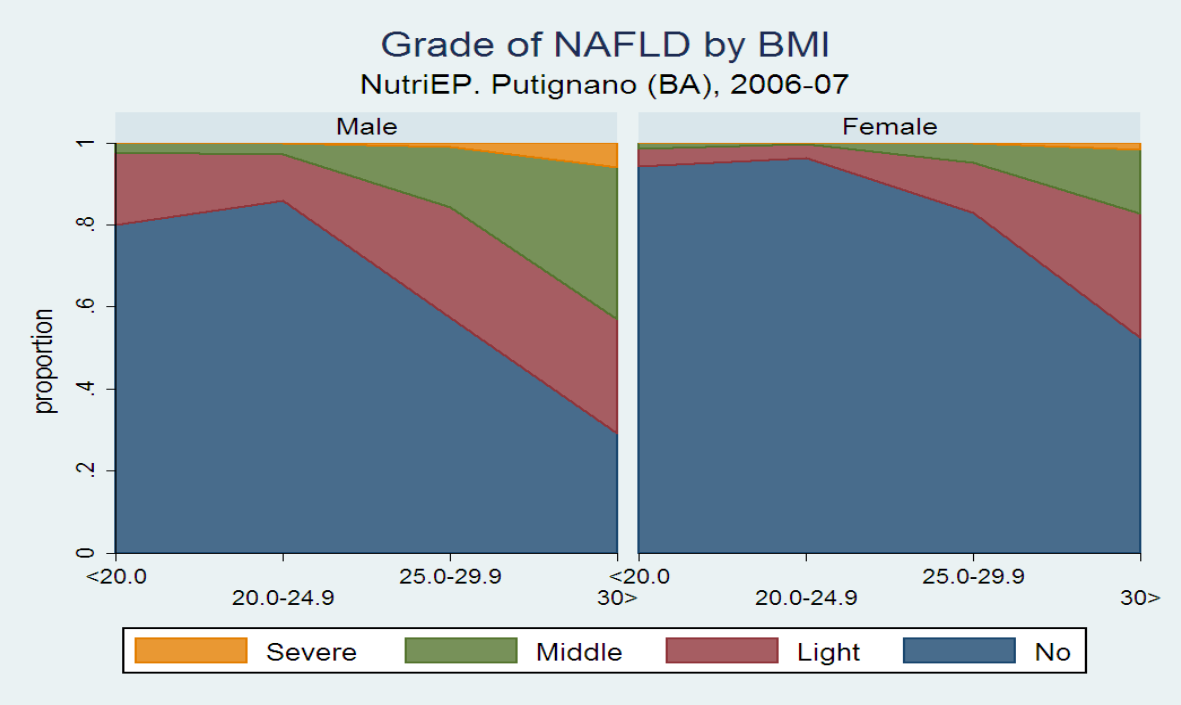
Results

- Descriptive Results
- Sequential Model
- Relationship between transitions and weights

Descriptive Results



Descriptive Results



The model

```
xi:seqlogit steato_grade i.StatCiv Etarecl glicemia
i.scalacat GOT GPT,
or
tree(0: 1 2 3, 1: 2 3, 2: 3)
ofinterest(BMI) over(Etarecl)
levels(0=0, 1=1.5, 2=4, 3=5.1) sd(0.25)
```

Sequential Model Fitting

```
                                                Number of obs   =       2176
                                                LR chi2(31)     =     764.05
Log likelihood = -1244.2097                     Prob > chi2     =     0.0000
```

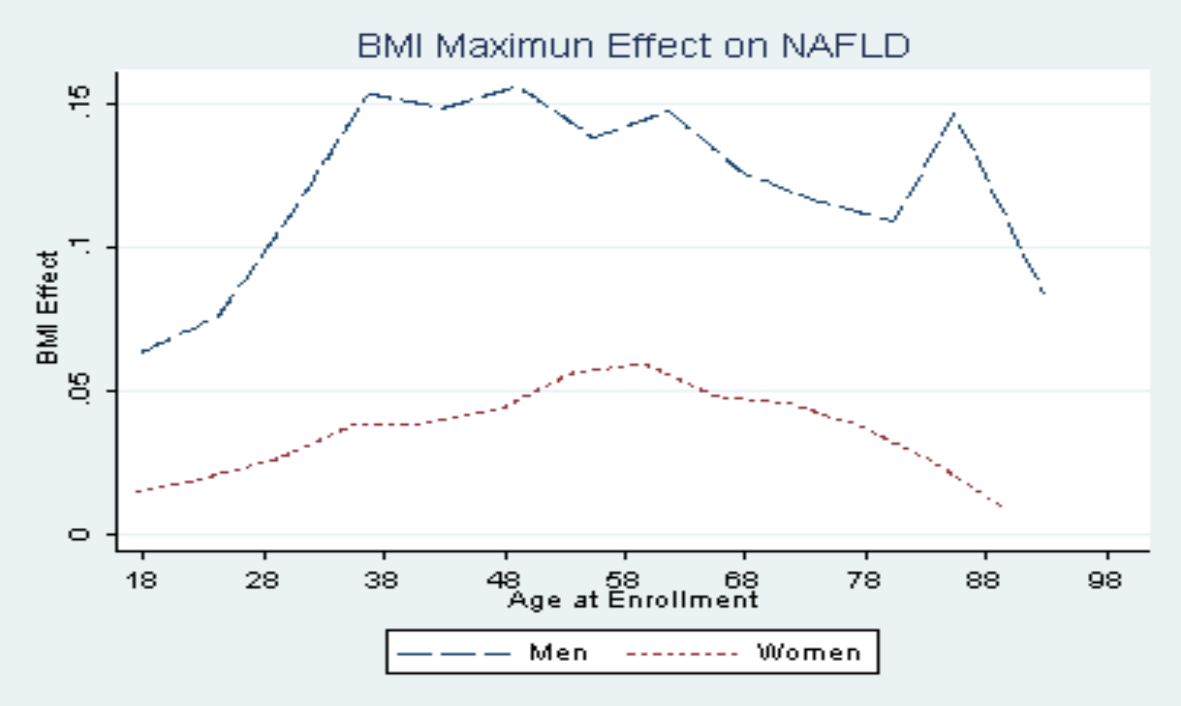
steato_grade	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_1_2_3v0						
_IStatCiv_2	1.857439	.4465768	2.58	0.010	1.159481	2.975539
_IStatCiv_3	1.888306	1.030708	1.16	0.244	.64782	5.504149
_IStatCiv_4	1.583159	.6337173	1.15	0.251	.7224373	3.469358
Etarecl	1.089943	.0343205	2.74	0.006	1.02471	1.159329
glicemia	1.013574	.0028864	4.73	0.000	1.007933	1.019248
_Iscalacat_1	3.089764	1.880696	1.85	0.064	.9371534	10.18685
_Iscalacat_2	1.26e-06	.0029656	-0.01	0.995	0	.
GOT	.9537327	.0195509	-2.31	0.021	.9161731	.992832
GPT	1.097814	.0123173	8.32	0.000	1.073937	1.122223
BMI	1.443293	.0866198	6.11	0.000	1.283126	1.623453
_BMI_X_Etarecl	.9972593	.001123	-2.44	0.015	.9950606	.9994628
_cons	4.93e-07	8.18e-08	-8.76	0.000	1.91e-08	.0000127

Syntax for predict

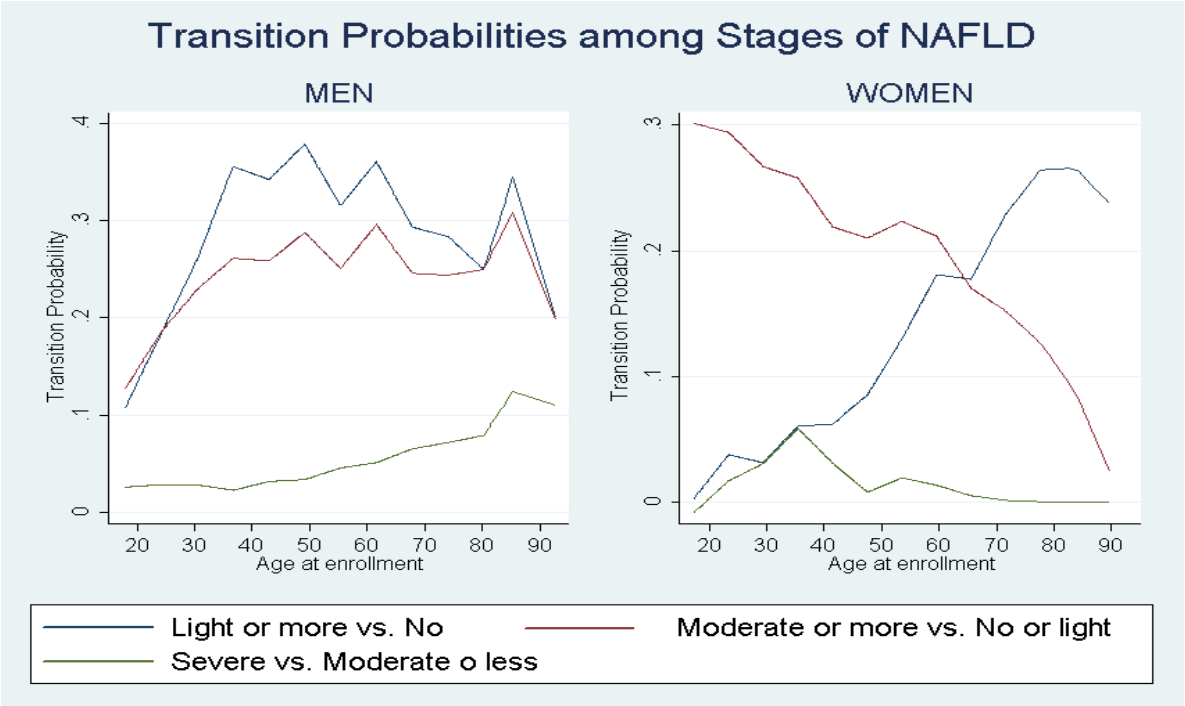
```
predict [type] newvar [if] [in] [, statistic outcome(#) transition(#)
        choice(#) equation(#) levels(levellist) ]
```

statistic	Description
xb	xb, fitted values
stdp	standard error of the prediction
trpr	probability of passing transition
tratrisk	proportion of respondents at risk of passing transition
trvar	variance of the indicator variable indicating whether or not the respondent passed the transition
trgain	difference in expected highest achieved level between those that pass the transition and those that do not
trweight	weight assigned to transition
pr	probability that an outcome is the highest achieved outcome.
y	expected highest achieved level
effect	Effect of variable of interest on expected highest achieved level. This variable is specified in the ofinterest() option in seqlogit. Interactions with the variables specified in the over() option of seqlogit are automatically taken into account.
residuals	difference between highest achieved level and expected highest achieved level.
score	first derivative of the log likelihood with respect to the linear predictor.

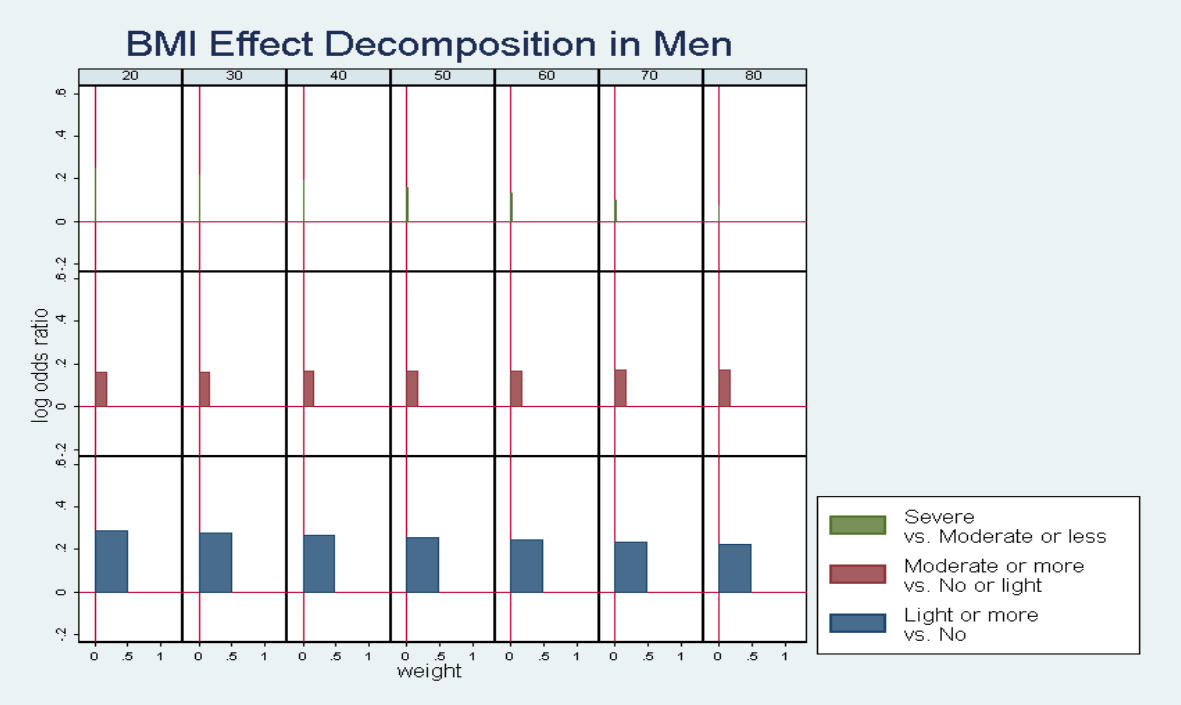
Relationship between transitions and weights



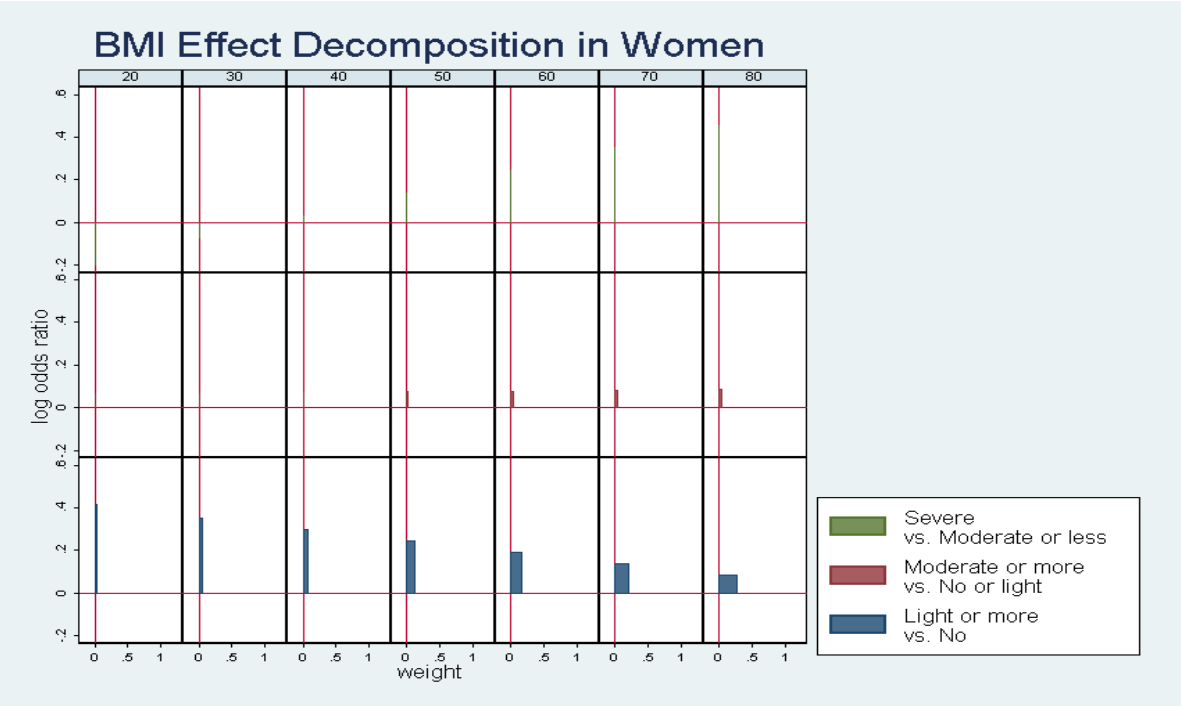
Relationship between transitions and weights



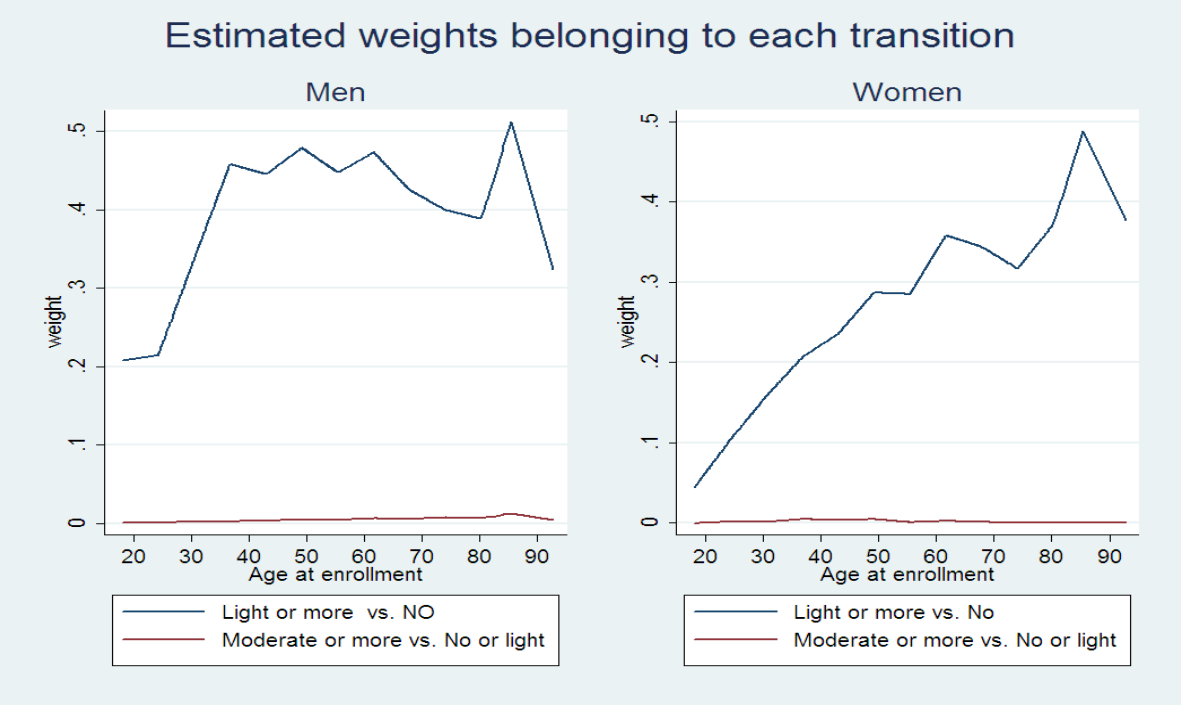
Relationship between transitions and weights



Relationship between transitions and weights



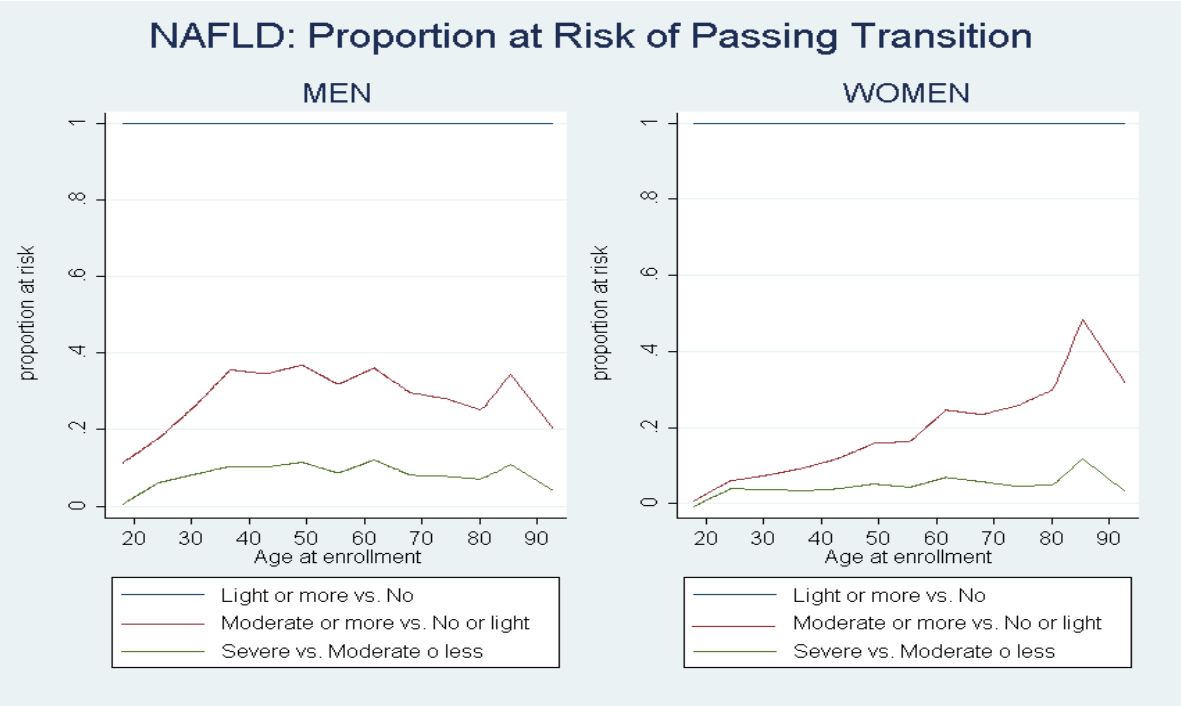
Relationship between transitions and weights



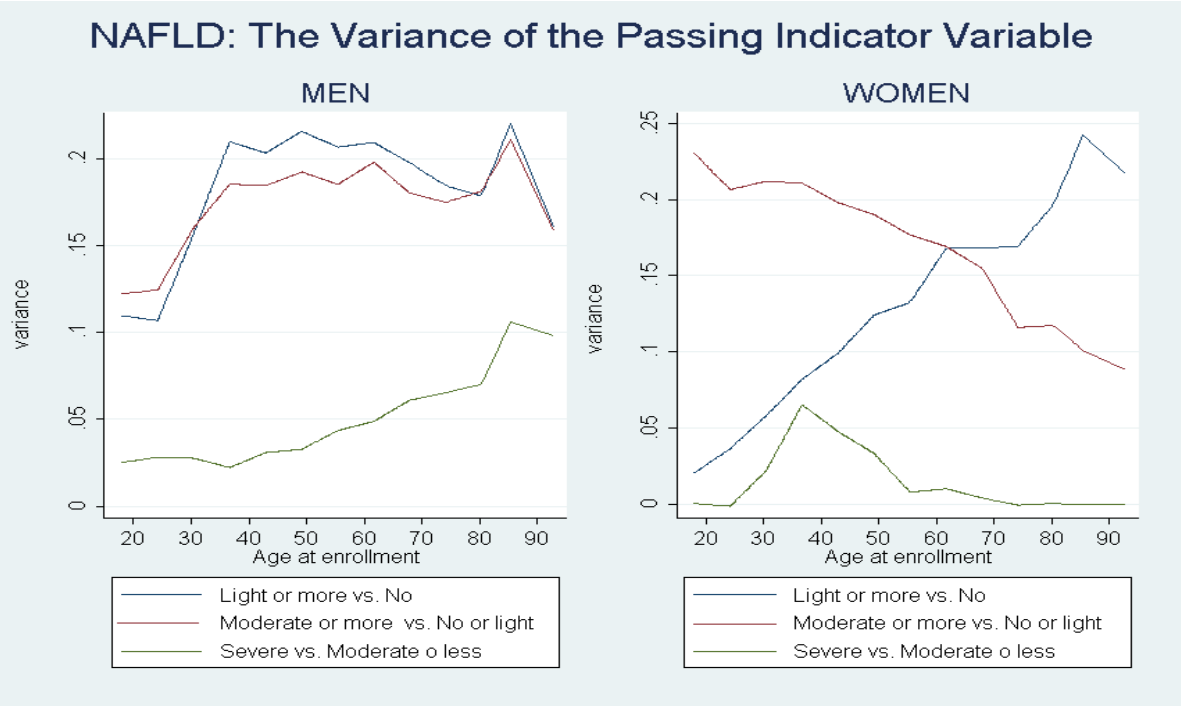
The weights are the product of three components:

- The proportion of people at risk at each transition
- The closeness to 50% of the proportion of people passing (variance)
- The difference in the expected severity of NAFLD between those passing and those failing a transition

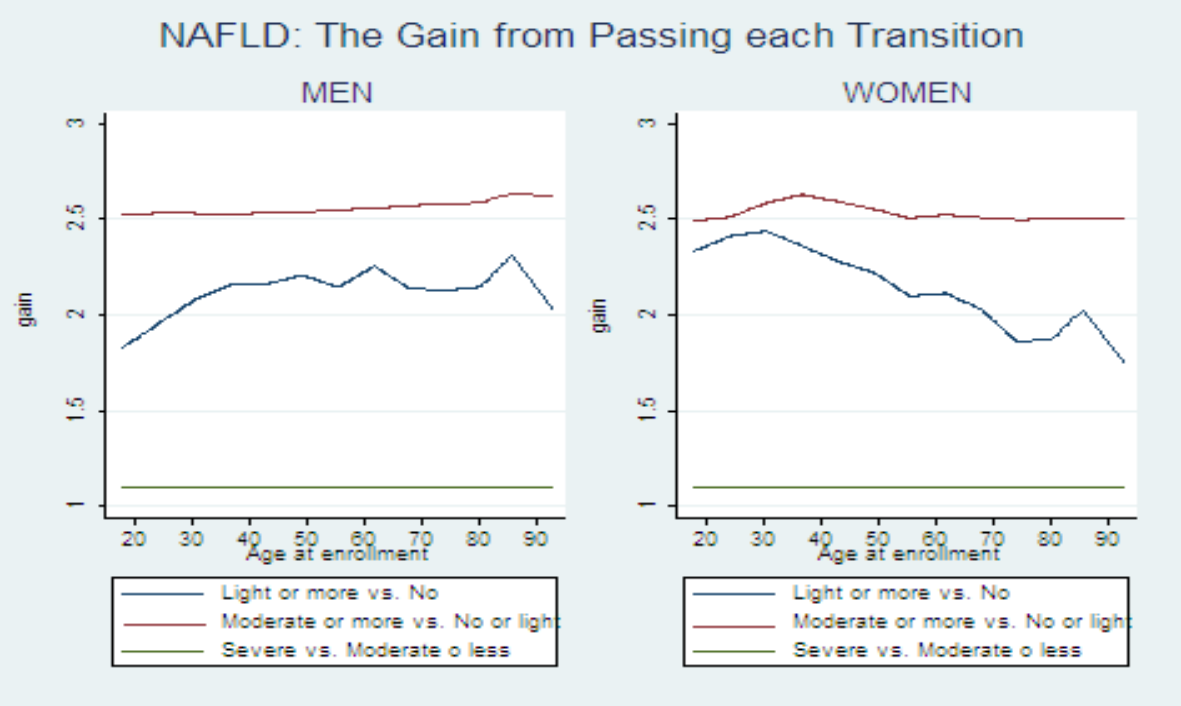
Relationship between transitions and weights



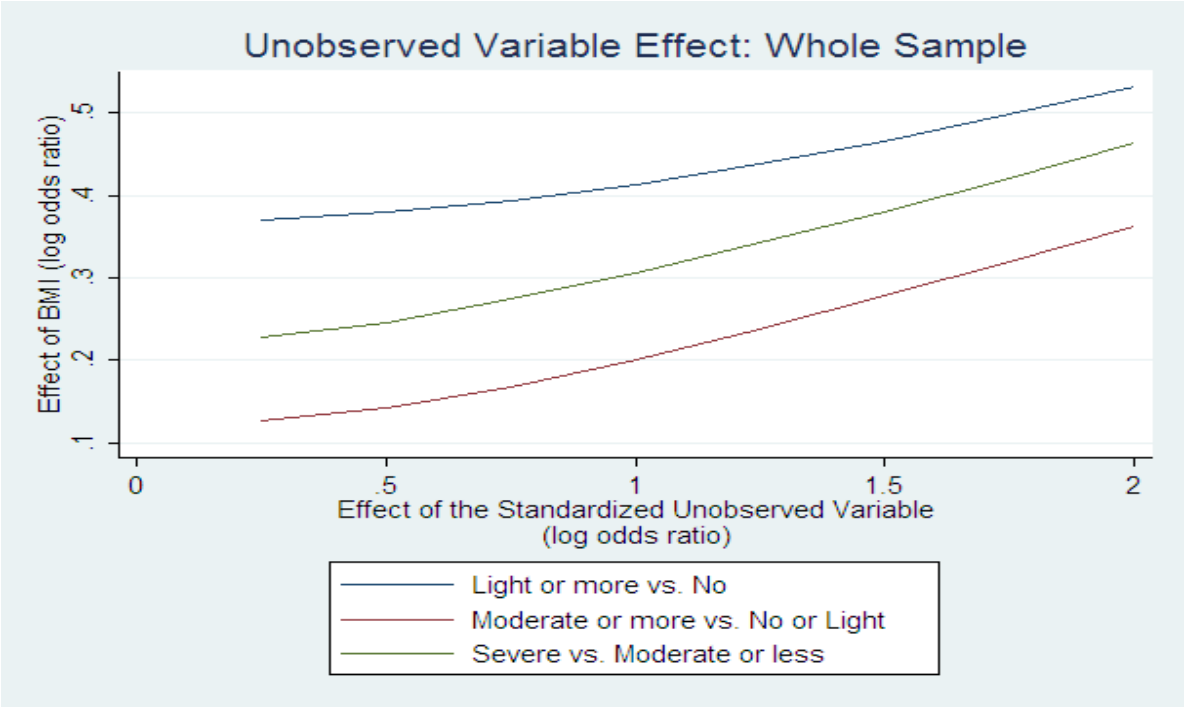
Relationship between transitions and weights



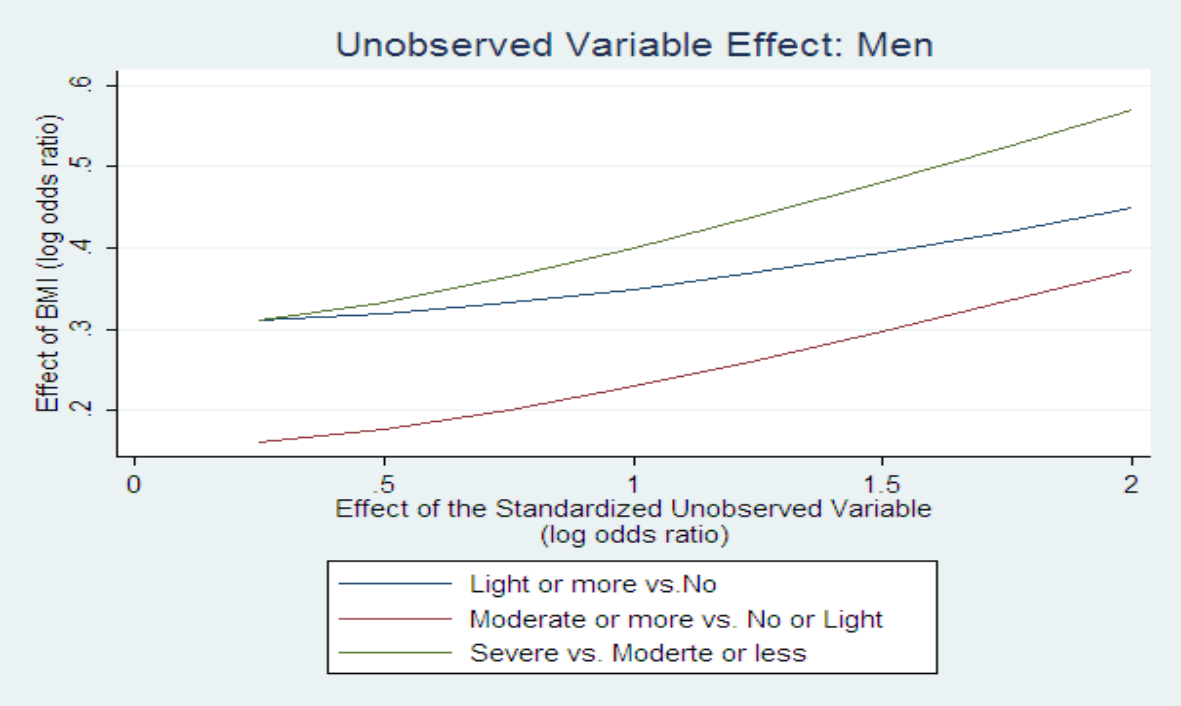
Relationship between transitions and weights



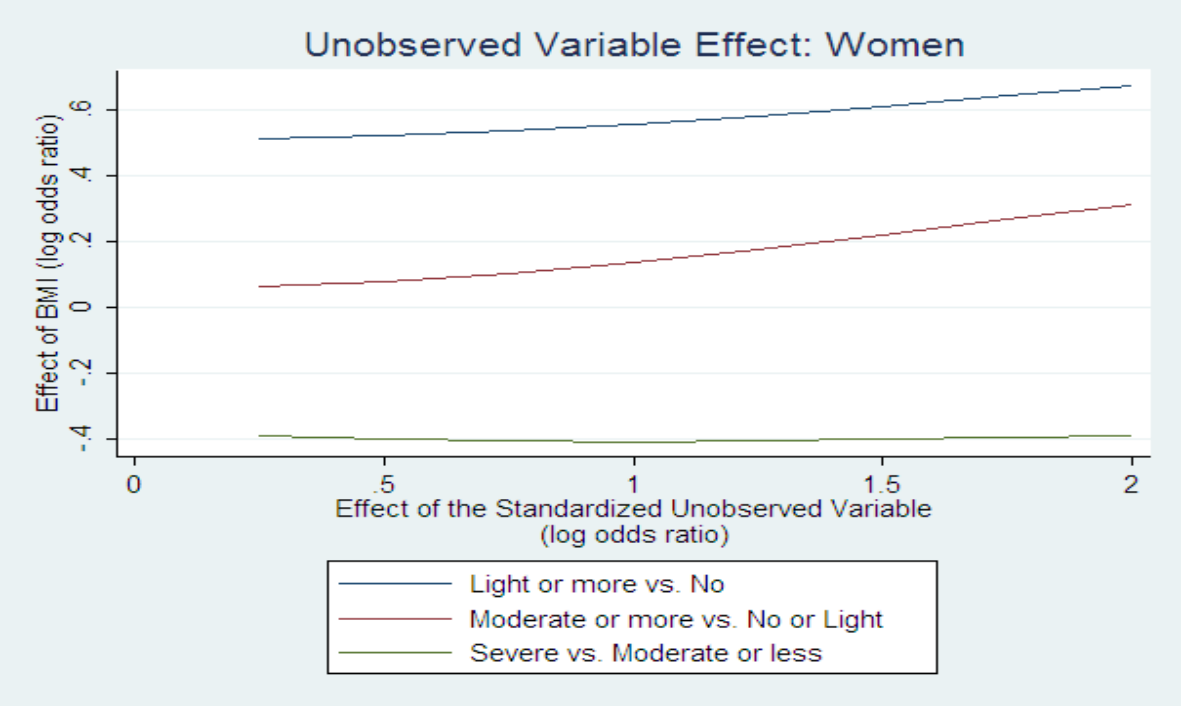
Sensitivity Analysis



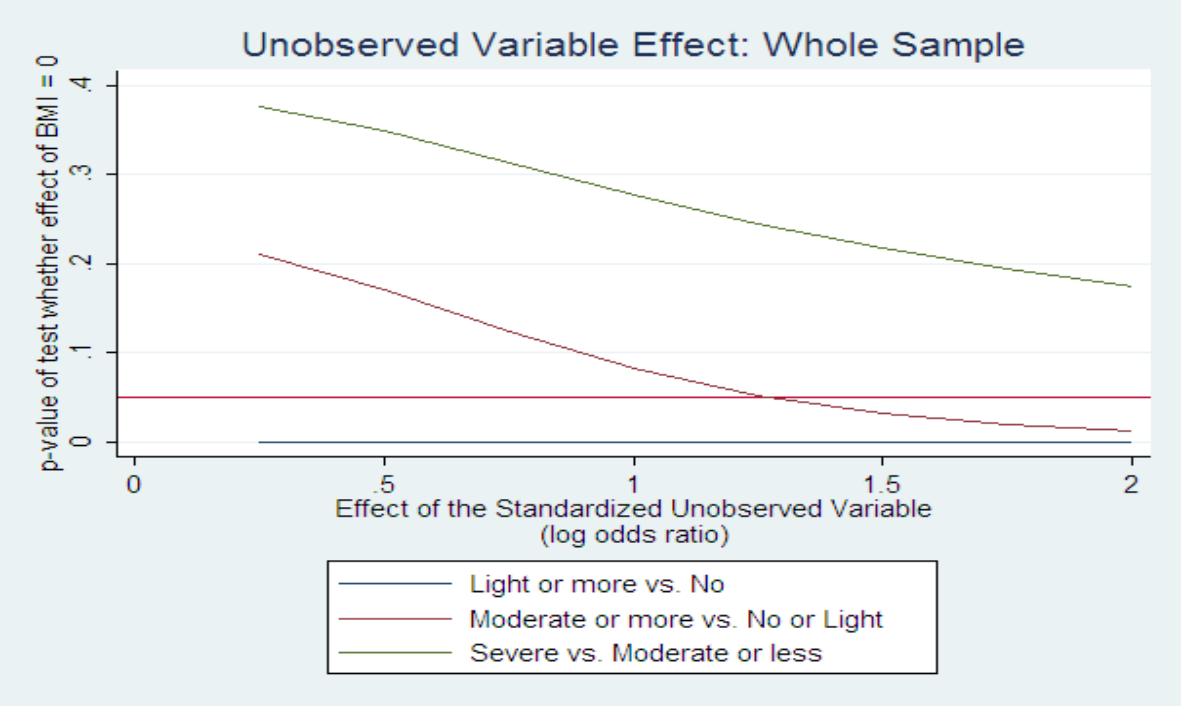
Sensitivity Analysis



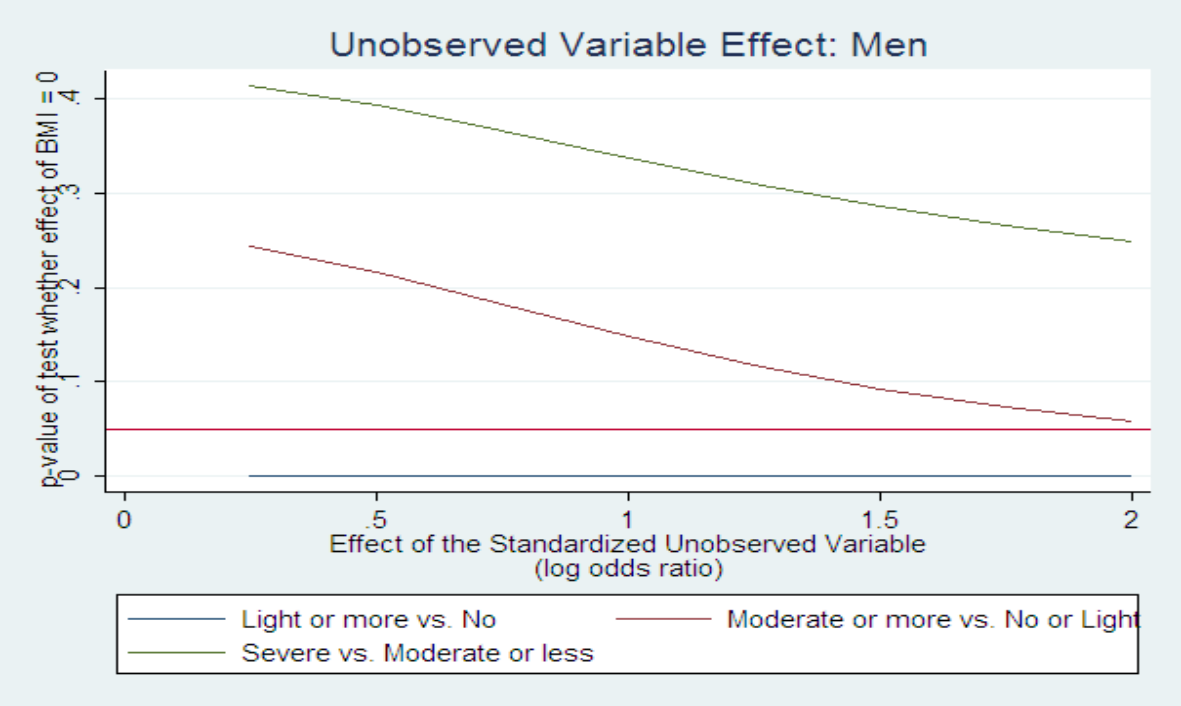
Sensitivity Analysis



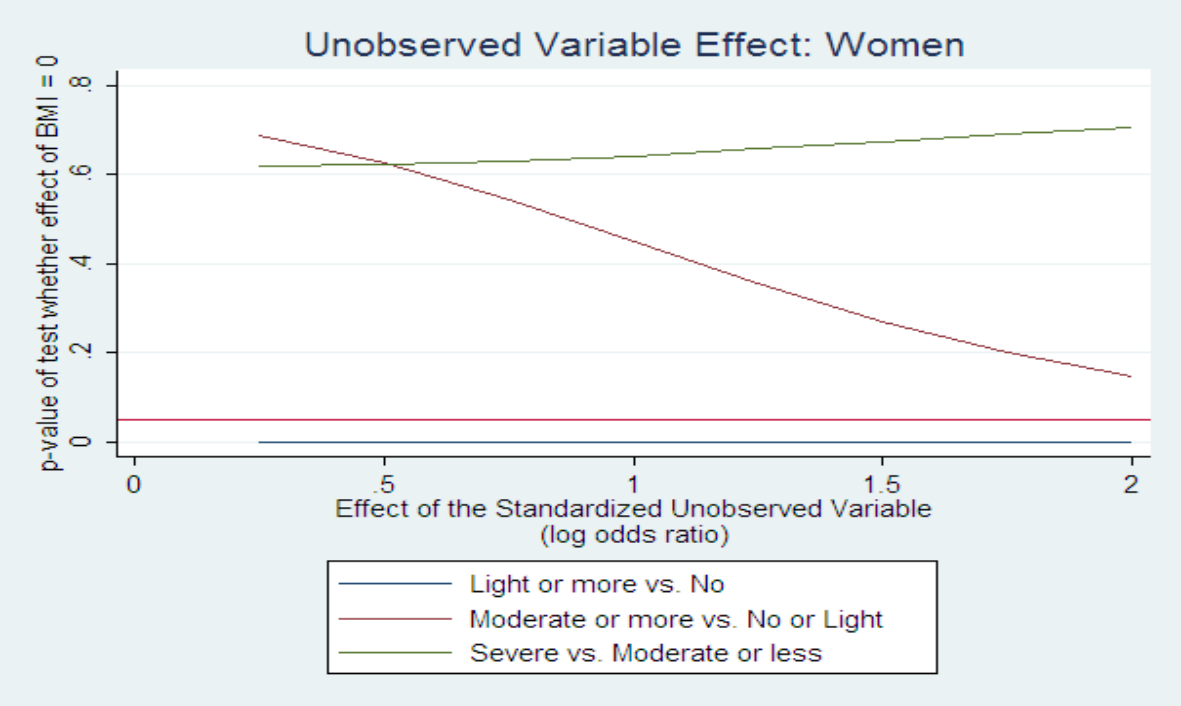
Sensitivity Analysis



Sensitivity Analysis



Sensitivity Analysis



Conclusions

- -seqlogit- is an usefull tool to explore transitions among different stages of a number of situations
- It's an user-friendly command
- It permits to perform a sensitivity analysis

Bibliography

- Fox, John 1997 Applied Regression Analysis, Linear Models, and Related Methods. Thousand Oaks: Sage.
- Maddala, G.S. 1983 Limited Dependent and Qualitative Variables in Econometrics Cambridge: Cambridge University Press.
- Mare, Robert D. 1981 “Change and Stability in educational Stratification” American Sociological Review, 46(1), p.p. 72-87.
- *http : //www.maartenbuis.nl/dissertation/chap_6.pdf*
- *http : //www.maartenbuis.nl/dissertation/chap_7.pdf*