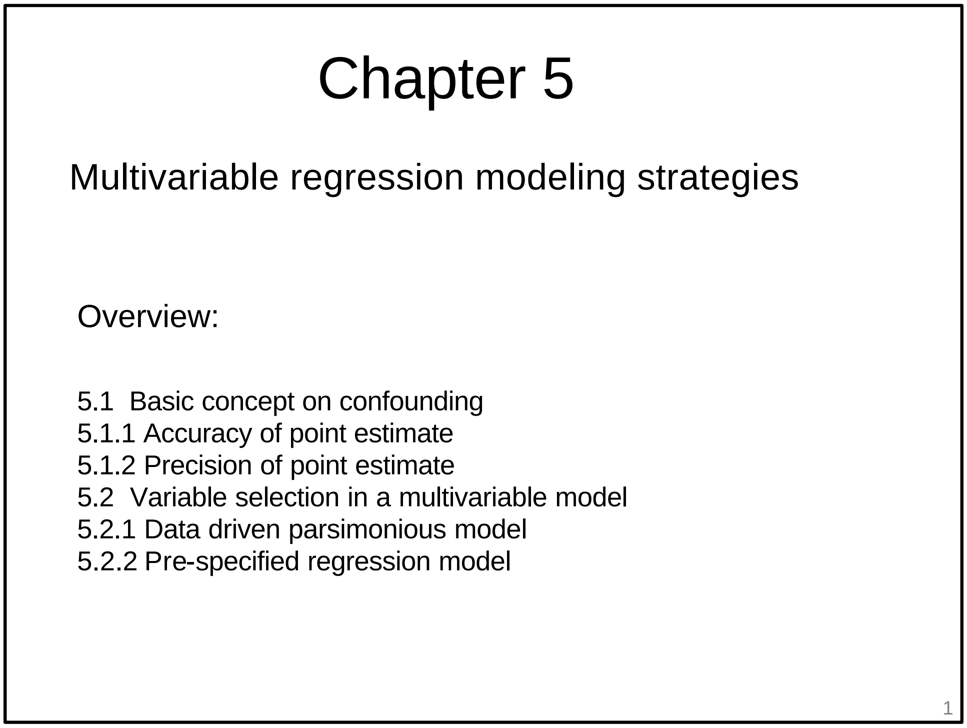
Master in Clinical Bio-Medical Science Program
Biostatistics II, July 2006
Ayumi Shintani, Ph.D., M.P.H.

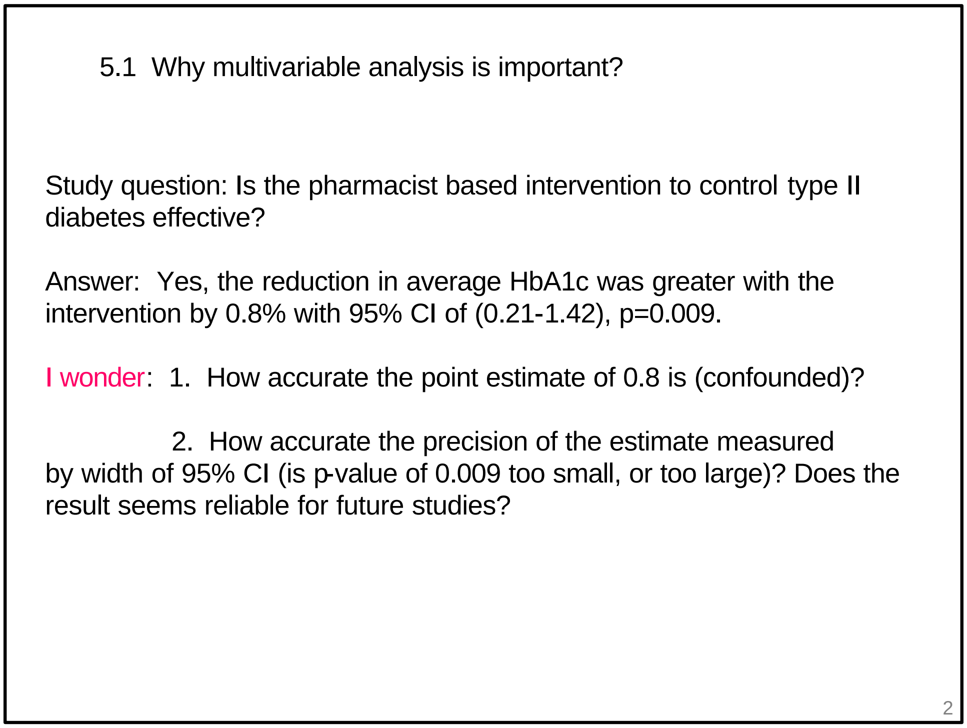


Chapter 5

Multivariable regression modeling strategies

Overview:

5.1 Basic concept on confounding
5.1.1 Accuracy of point estimate
5.1.2 Precision of point estimate
5.2 Variable selection in a multivariable model
5.2.1 Data driven parsimonious model
5.2.2 Pre-specified regression model



5.1 Why multivariable analysis is important?

Study question: Is the pharmacist based intervention to control type II diabetes effective?

Answer: Yes, the reduction in average HbA1c was greater with the intervention by 0.8% with 95% CI of (0.21-1.42), p=0.009.

I wonder: 1. How accurate the point estimate of 0.8 is (confounded)?

2. How accurate the precision of the estimate measured by width of 95% CI (is p-value of 0.009 too small, or too large)? Does the result seems reliable for future studies?

5.1.1. How accurate the point estimate of 0.8 is?

Point estimate of the effect of treatment can often over or under estimate a true effect of treatment by existence of a confounding factor. When such confounding factor is not considered in a study design, it must be controlled in statistical analysis.

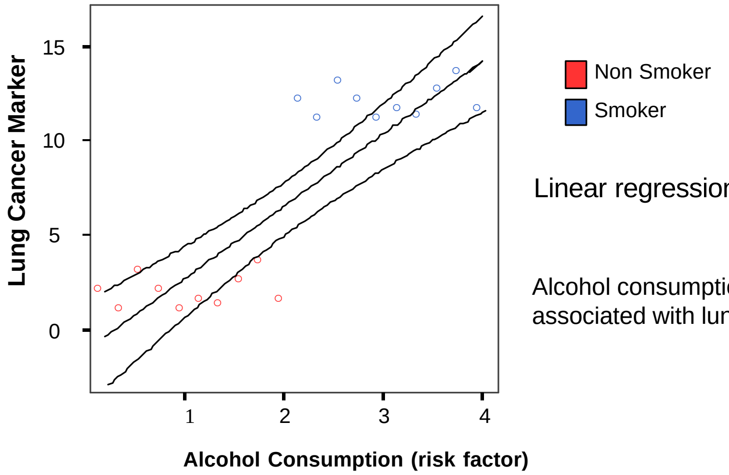
RCT – randomization prevents confounders:

i.e., in order to be a confounder, the extraneous factor must be associated with both outcome and exposure. Through randomization, treatment assignment tends to be balanced to both observed and unobserved extraneous factors. Thus estimated effect of treatment from unadjusted analysis is probably accurate (unbiased).

Observational studies – without randomization, treatment effect often be biased by the extraneous factor which is associated with an exposure of interest, thus adjusted analysis almost always “must” be used.

Graphical Presentation of Confounder (1):Assessing confounder by stratified analysis

Ignoring Smoking (Confounder)

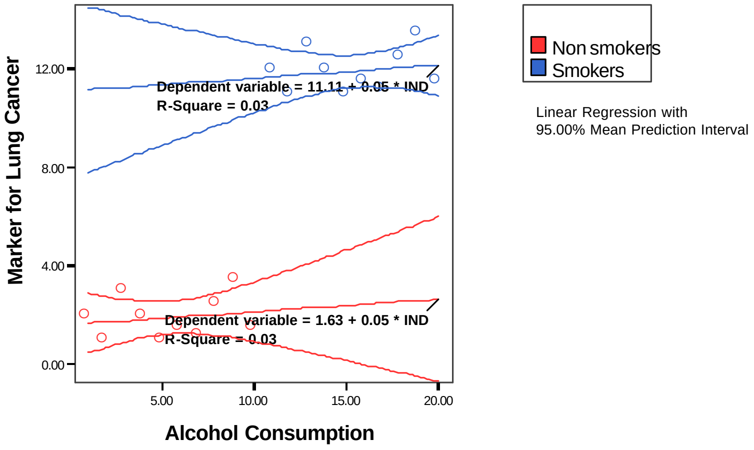


Linear regression with 95% CI

Alcohol consumption seem to be associated with lung cancer.

Graphical Presentation of Confounder (1):Assessing confounder by stratified analysis

Stratified by Confounder



Mathematical assessment of confounding

Remember,

Adjusted effect of X for Conf

Adjusted: $Y = a + b_1 X + b_2 Conf$

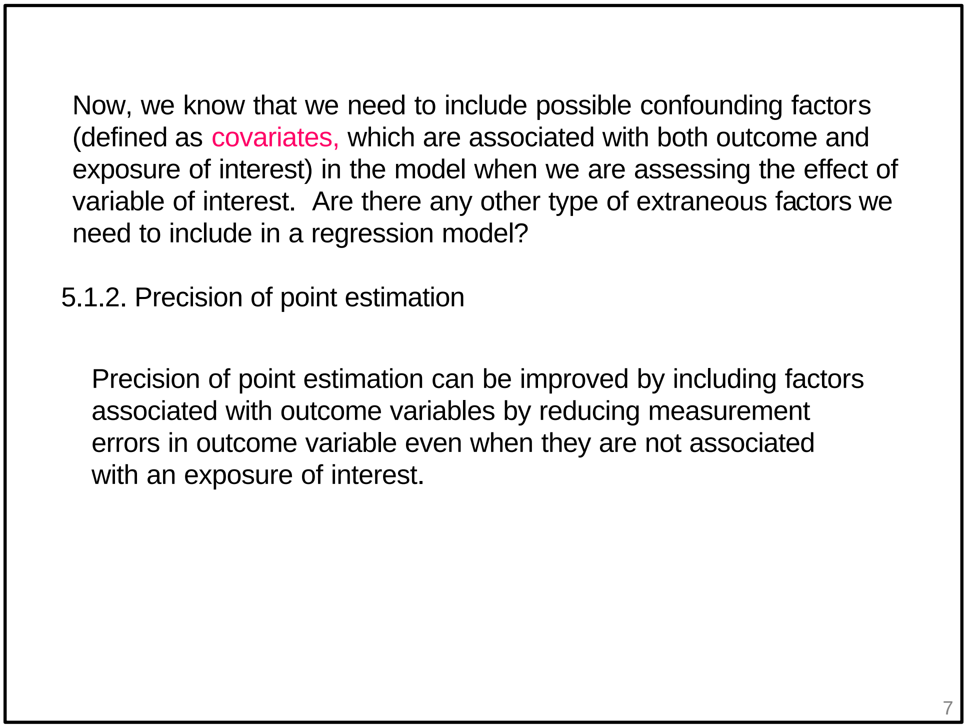
Un-adjusted: $Y = a + b'_1 X$

Unadjusted effect

$b_1 \neq b'_1$ Evidence of C being a confounder

$b_1 = b'_1$ No evidence of C being a confounder

Many people define confounder if $(b_1 - b'_1)/b'_1$ >0.10, 0.15 or 0.2 regardless of p-value

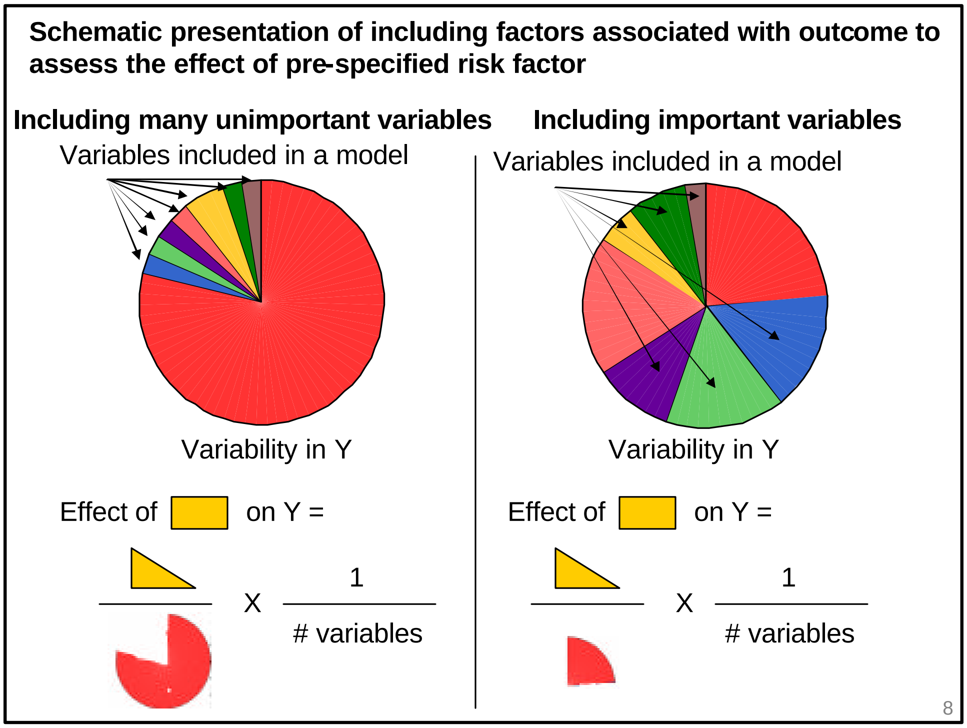


Now, we know that we need to include possible confounding factors (defined as covariates, which are associated with both outcome and exposure of interest) in the model when we are assessing the effect of variable of interest. Are there any other type of extraneous factors we need to include in a regression model?

5.1.2. Precision of point estimation

Precision of point estimation can be improved by including factors associated with outcome variables by reducing measurement errors in outcome variable even when they are not associated with an exposure of interest.

7



Schematic presentation of including factors associated with outcome to assess the effect of pre-specified risk factor

Including many unimportant variables

Variables included in a model

Variability in Y

Effect of 🟨 on Y =

× 1 / # variables

Including important variables

Variables included in a model

Variability in Y

Effect of 🟨 on Y =

× 1 / # variables

8

5.2 Variable selection in a multivariable model

Result of including confounding variable in a model

When Factor A is a confounder to the association between outcome and intervention, including Factor A will change parameter estimate of the intervention compared with that of unadjusted model.

Result of including risk factor of outcome in a model

If Factor A is a risk factor of the outcome variable, including Factor A can remove variability (measurement error) of data, thus standard error of the estimate for the intervention effect tends to be reduced, resulting a smaller p-value for the intervention effect (see the schematic explanation in the previous page).

Result of including neither risk factor or confounder

Including variables which is neither associated with exposure nor outcome variable will lead to loss of statistical power without a gain (see the schematic explanation in the previous page).

Exercise: Select variables to include in a linear regression model to assess risk factor of post traumatic stress disorder (PTSD) among 43 ICU survivors. Data were collected for the following variables:

Age
Gender
Race
Apache II severity of illness score
SOFA (score of organ function)
Baseline dementia score
Hearing difficulty
Vision difficulty

Admission diagnosis of sepsis
Presence of depression
Alcohol abuse
Drug overdose
Ability of daily living (ADL)
ICU days of delirium
ICU days of coma
ICU length of stay in days
Days of mechanical ventilation
ICU use of sedative drug (lorazapam)

Variable selection for a multivariable model (model building): determining how many variable?

Guideline for the maximum number of independent variables (degree of freedom) to be included in a multivariable model.

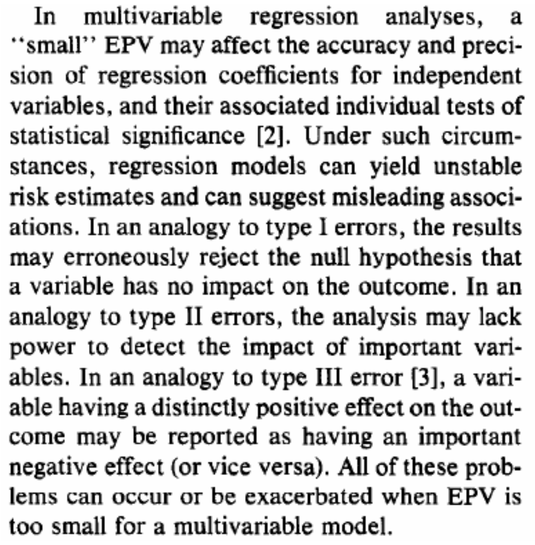
Linear regression	# patients (samples) / 15 (10-20)
Logistic regression	Min(#events, #non-events) / 15 (10-20)
Cox regression	#events / 15 (10-20)
Proportional odds logistic regression	$n - \frac{1}{n^2}\sum_{i=1}^{k} n_i^3$ / 15 (10-20)

K: number of categories, n: total sample size, n_i: sample size in each category

References:
* Harrell FE, Jr. Regression Modeling Strategies. Springer Verlag. (2001).
* Peduzzi P et al. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996 Dec;49(12):1373-9.
* Peduzzi P et al. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol. 1995 Dec;48(12):1503-10.

11

Problems with over-fitted model:



In multivariable regression analyses, a "small" EPV may affect the accuracy and precision of regression coefficients for independent variables, and their associated individual tests of statistical significance [2]. Under such circumstances, regression models can yield unstable risk estimates and can suggest misleading associations. In an analogy to type I errors, the results may erroneously reject the null hypothesis that a variable has no impact on the outcome. In an analogy to type II errors, the analysis may lack power to detect the impact of important variables. In an analogy to type III error [3], a variable having a distinctly positive effect on the outcome may be reported as having an important negative effect (or vice versa). All of these problems can occur or be exacerbated when EPV is too small for a multivariable model.

John Concato, Peter Peduzzi, Theodore R. Holford and Alvan R. Feinstein. Importance of events per independent variable in proportional hazards analysis I. Journal of Clinical Epidemiology. Vol 48 (12) December 1995, Pages 1495-1501

12

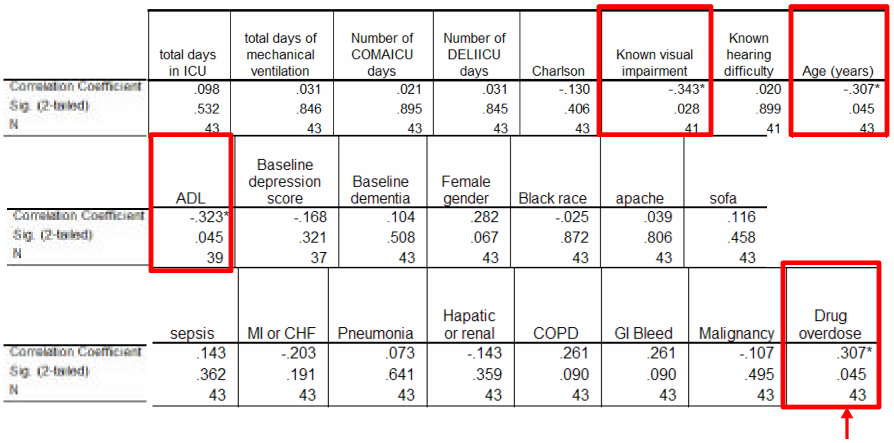
How many variables to include in the PTSD analysis?

In order to prevent from over-fitting, based on the previous page, when we use linear regression to fit this data, we are able to include only up to 4 variables (43/10).

Which variables to include?

We want to include confounders of the association between any identified risk factors and PTSD to obtain unbiased estimate of the effect of risk factors, and also factors associated with PTSD to reduce measurement errors. Since confounding factor requires to be associated with both risk factor and outcome (not to be a confounder if not associated with outcome), thus, we include factors which are associated with outcome for simplicity.

Spearman's correlation coefficients with PTSD



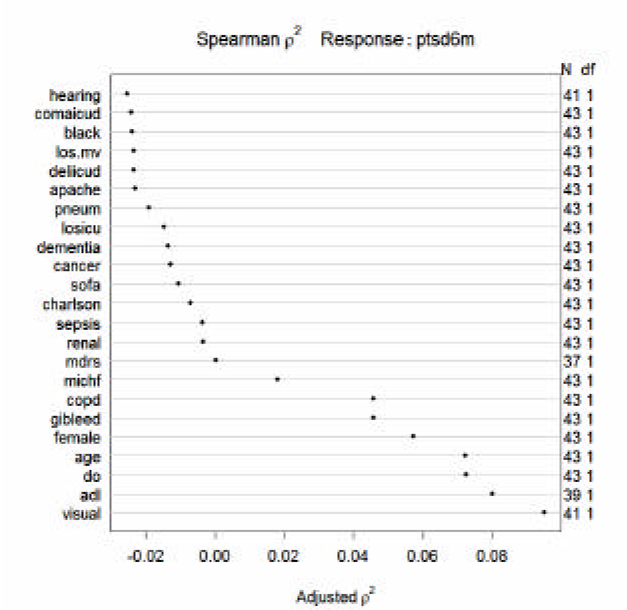
P=0.03 (Man-Whitney U)

	total days in ICU	total days of mechanical ventilation	Number of COMAICU days	Number of DELIICU days	Charlson	Known visual impairment	Known hearing difficulty	Age (years)
Correlation Coefficient	.098	.031	.021	.031	-.130	-.343*	.020	-.307*
Sig. (2-tailed)	.532	.846	.895	.845	.406	.028	.899	.045
N	43	43	43	43	43	41	41	43

	ADL	Baseline depression score	Baseline dementia	Female gender	Black race	apache	sofa
Correlation Coefficient	-.323*	-.168	.104	.282	-.025	.039	.116
Sig. (2-tailed)	.045	.321	.508	.067	.872	.806	.458
N	39	37	43	43	43	43	43

	sepsis	MI or CHF	Pneumonia	Hapatic or renal	COPD	GI Bleed	Malignancy	Drug overdose
Correlation Coefficient	.143	-.203	.073	-.143	.261	.261	-.107	.307*
Sig. (2-tailed)	.362	.191	.641	.359	.090	.090	.495	.045
N	43	43	43	43	43	43	43	43

P=0.046
Man-Whitney U



We may choose age, drug overdose, ADL, and visual difficulty to include

15

Alternative method to present association between risk factor vs PTSD

```
Descriptive Statistics by (ptsd6m > 21)

+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|                                     |N |FALSE                  |TRUE                   |Combined               |  Test                          |
|                                     |  |(N=23)                 |(N=20)                 |(N=43)                 |Statistic                       |
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|total days in ICU                    |43|    5.0/10.0/14.0      |    6.0/ 9.5/13.0      |    5.5/10.0/13.0      |    F=0 d.f.=1,41 P=0.98        |
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|total days of mechanical ventilation|43|    2.0/ 7.0/12.5      |    4.0/ 5.0/ 7.5      |    3.5/ 5.0/11.5      |   F=0.09 d.f.=1,41 P=0.763     |
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Number of COMAICU days               |43|   0.50/1.00/2.00      |   0.00/1.00/4.25      |   0.00/1.00/3.00      |   F=0.03 d.f.=1,41 P=0.873     |
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Number of DELIICU days               |43|        1/2/3          |        1/2/3          |        1/2/3          |    F=0 d.f.=1,41 P=0.97        |
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Charlson                             |43|   2.00/3.00/4.00      |   1.75/3.50/6.00      |   2.00/3.00/5.00      |   F=0.65 d.f.=1,41 P=0.424     |
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Known visual impairment              |41|      73% (16)         |      53% (10)         |      63% (26)         |Chi- square=1.77 d.f.=1 P=0.183|
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Known hearing difficulty             |41|       9% ( 2)         |      21% ( 4)         |      15% ( 6)         | Chi -square=1.17 d.f.=1 P=0.28|
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Age (years)                          |43|   46.0/59.0/70.0      |   39.0/51.0/53.5      |   40.0/52.0/63.0      |   F=3.19 d.f.=1,41 P=0.0813    |
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|adl                                  |39|        0/0/0          |        0/0/0          |        0/0/0          |   F=1.82 d.f.=1,37 P=0.186     |
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Baseline depression score            |37|        0/0/0          |        0/0/0          |        0/0/0          |   F=0.68 d.f.=1,35 P=0.415     |
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Baseline dementia                    |43|      17% ( 4)         |      20% ( 4)         |      19% ( 8)         |Chi- square=0.05 d.f.=1 P=0.826|
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Female gender                        |43|      48% (11)         |      60% (12)         |      53% (23)         |Chi- square=0.64 d.f.=1 P=0.425|
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Black race                           |43|      17% ( 4)         |      15% ( 3)         |      16% ( 7)         |Chi- square=0.04 d.f.=1 P=0.832|
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|apache                               |43|  18.00/24.00/31.00|  22.75/26.50/30.25|  20.50/25.00/30.50|   F=0.55 d.f.=1,41 P=0.464     |
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|sofa                                 |43|    7.5/10.0/12.0      |    8.0/11.0/12.0      |    8.0/11.0/12.0      |   F=0.47 d.f.=1,41 P=0.497     |
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|sepsis                               |43|      35% ( 8)         |      50% (10)         |      42% (18)         |Chi- square=1.02 d.f.=1 P=0.313|
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|MI or CHF                            |43|      13% ( 3)         |       5% ( 1)         |       9% ( 4)         |Chi- square=0.82 d.f.=1 P=0.365|
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Pneumonia                            |43|      22% ( 5)         |      30% ( 6)         |      26% (11)         |Chi- square=0.38 d.f.=1 P=0.536|
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Hapatic or renal                     |43|      13% ( 3)         |      10% ( 2)         |      12% ( 5)         | Chi -square=0.1 d.f.=1 P=0.756|
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|copd                                 |43|        0% ( 0)        |        5% ( 1)        |        2% ( 1)        |Chi- square=1.18 d.f.=1 P=0.278|
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|GI Bleed                             |43|        0% ( 0)        |        5% ( 1)        |        2% ( 1)        |Chi- square=1.18 d.f.=1 P=0.278|
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Malignancy                           |43|       4% ( 1)         |       5% ( 1)         |       5% ( 2)         | Chi -square=0.01 d.f.=1 P=0.92|
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
|Drug overdose                        |43|        0% ( 0)        |       10% ( 2)        |        5% ( 2)        |Chi- square=2.41 d.f.=1 P=0.120|
+-------------------------------------+--+-----------------------+-----------------------+-----------------------+--------------------------------+
Warning messages:
```

16

Thus the final model based on the univariate analysis may include age, known visual impairment, ADL and drug overdose.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	34.431	6.527		5.275	.000	21.166	47.695
	Age (years)	-.159	.133	-.197	-1.195	.240	-.429	.111
	ADL	-1.301	.846	-.224	-1.539	.133	-3.020	.417
	Known visual impairment	-4.967	3.761	-.218	-1.321	.195	-12.610	2.676
	Drug overdose	13.279	7.325	.264	1.813	.079	-1.607	28.165

a. Dependent Variable: PTSD score at 6 months post hosp discharge

17

Data driven parsimonious model

Traditionally, many believed parsimonious modeling strategies, model which includes fewer number of variables to explain greater variability in outcome variable is better, which in many cases, resulting in excluding in-significant variables from a regression model.

Popular approach in parsimonious model building includes

(1) Univaraite selection – include variables which are significant at univariate analysis
(2) Computer automated computer procedure to select a set of variables which are plausible to explain variability in outcome variable.
(3) Combination of the above (1) and (2): include variables selected in (1) into (2) to further select variables

18

Computer automated selection procedures

a) Forward Selection

Assess all simple linear regression with each independent variable in a model, then pick variables with the smallest p-value if p value is less than a cutoff (i.e., 0.15). By keeping the variable selected above, repeat the procedure for the remaining variables until the model include only variables with p values less than the cutoff level.

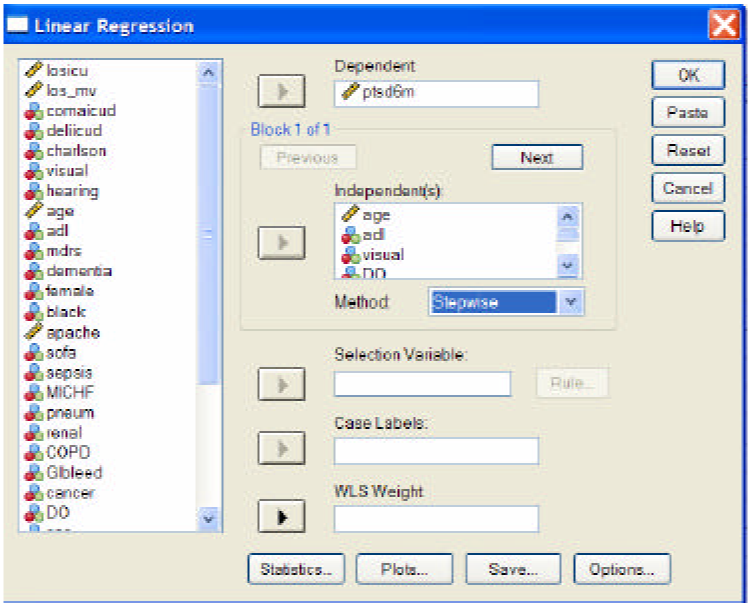
b) Backward Selection

This method is similar to the forward method except that we start with all the variables and eliminate the variable with the least significance. The data is refit with the remaining variables and the process is repeated until all remaining variables have a significance level below some threshold.

c) Stepwise Selection

This method is like the forward method except that at each step, previously selected variables whose significance has dropped below some threshold are dropped from the model.

19

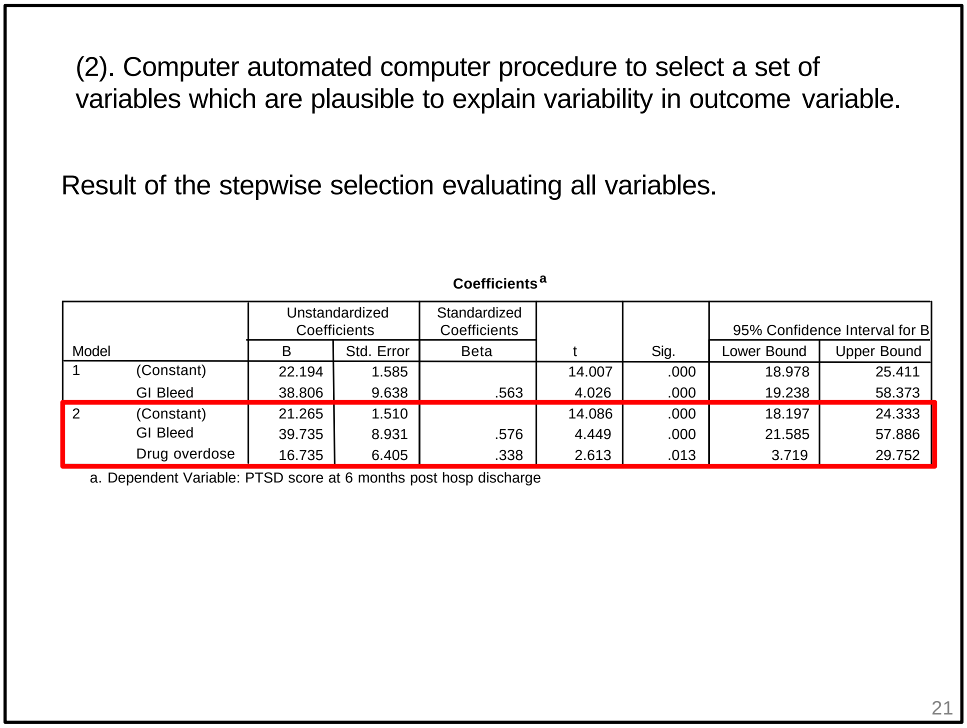


Dependent: ptsd6m

Independent(s): age, adl, visual, DO

Method: Stepwise

20



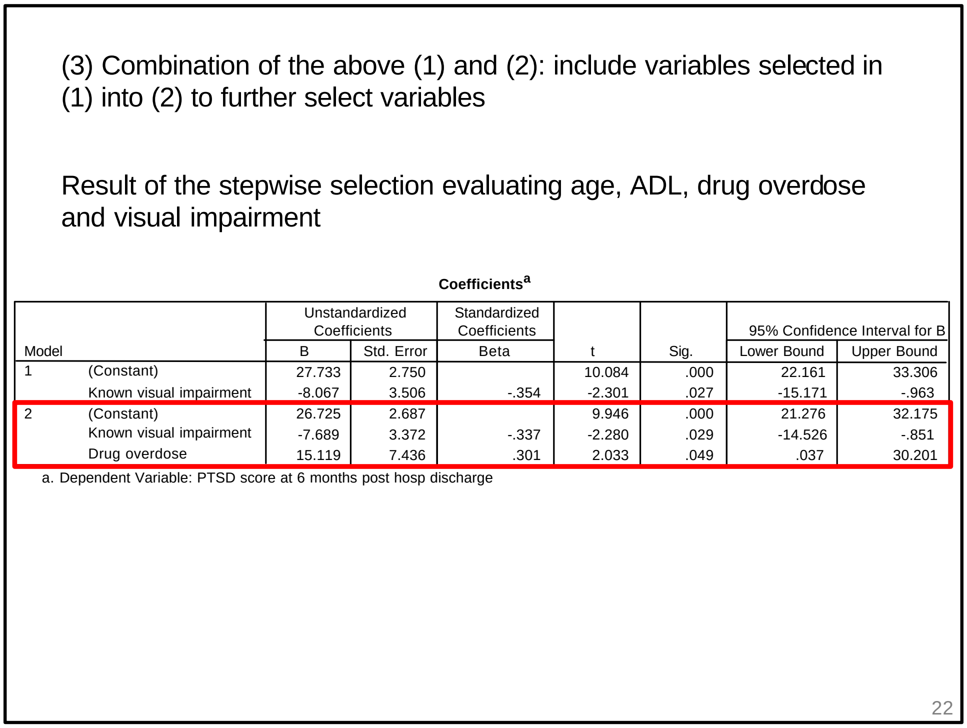
(2). Computer automated computer procedure to select a set of variables which are plausible to explain variability in outcome variable.

Result of the stepwise selection evaluating all variables.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	22.194	1.585		14.007	.000	18.978	25.411
	GI Bleed	38.806	9.638	.563	4.026	.000	19.238	58.373
2	(Constant)	21.265	1.510		14.086	.000	18.197	24.333
	GI Bleed	39.735	8.931	.576	4.449	.000	21.585	57.886
	Drug overdose	16.735	6.405	.338	2.613	.013	3.719	29.752

a. Dependent Variable: PTSD score at 6 months post hosp discharge



(3) Combination of the above (1) and (2): include variables selected in (1) into (2) to further select variables

Result of the stepwise selection evaluating age, ADL, drug overdose and visual impairment

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	27.733	2.750		10.084	.000	22.161	33.306
	Known visual impairment	-8.067	3.506	-.354	-2.301	.027	-15.171	-.963
2	(Constant)	26.725	2.687		9.946	.000	21.276	32.175
	Known visual impairment	-7.689	3.372	-.337	-2.280	.029	-14.526	-.851
	Drug overdose	15.119	7.436	.301	2.033	.049	.037	30.201

a. Dependent Variable: PTSD score at 6 months post hosp discharge

However exhaustive data searching for a parsimonious model including univariate selection and computer automated model selection has recently been heavily criticized for it inflates type I error (over-fitting). Because this is essentially the same as fitting many regressions which generate many p-values, therefore, the final model chosen by these procedures usually make standard error smaller than it should be.

References:
Harrell, Regression Modeling Strategies.http://www.cmh.edu/stats/faq/faq12.asp
Altman, D. G. and Andersen, P. K, Bootstrap investigation of the stability of a Cox regression model. Statistics in Medicine (1989) vol8:771-783
Derksen, S. and Keselman, H. J. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. British Journal of Mathematical and Statistical Psychology (1992). Vol45: 265-282

5.2.2 Pre-specified regression model

My recommendation of selecting variable in a regression is:

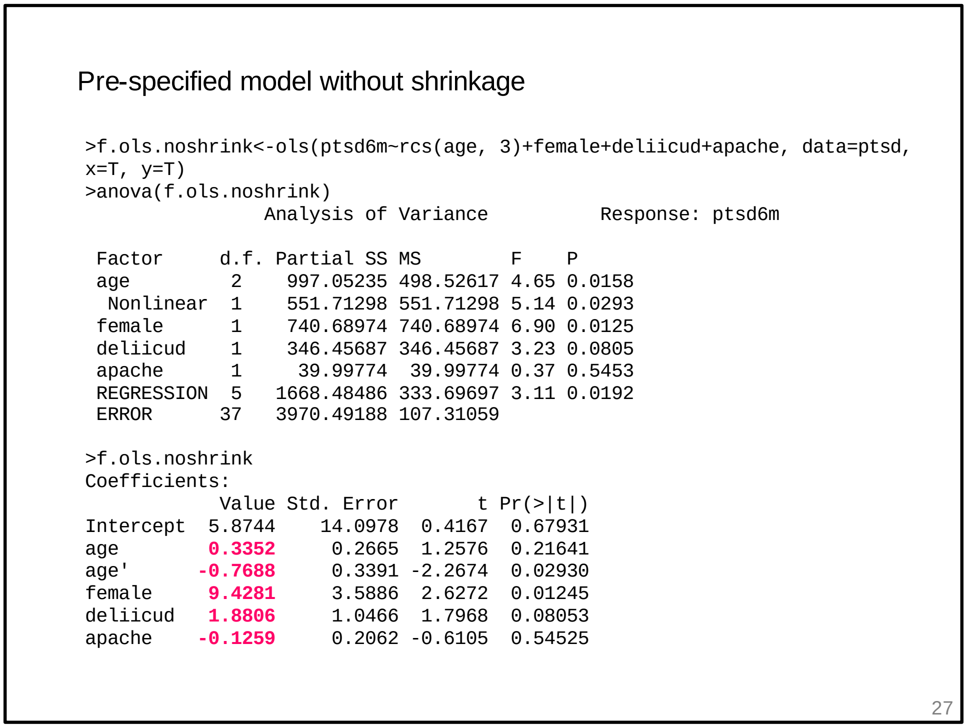
A priori (not looking at data) choose potential risk factors to outcome variable within the allowable limit determined by the general rule (You should not exclude insignificant variables!).

In fact, allowable number of variables (degree of freedom) to be included, will be affected by the number of dummies, and whether you want to asses non-linearity for continuous variables. You need minimum of 2 or 3 degree of freedoms to fit 1 non-linear continuous variable. Here I in fact chose age, gender, delirium days and Apace score to be included based on a prior belief. We are also interested in non-linear effect of age (we can use generalized Spearman's analysis to decide which variable to allow non-linear effect). SPSS cannot do non-linear associations, so we used R-software (total number of degree of freedom was (1+1+1+3=6) which indicates a slight over-fitting for the allowable number was 43/10=4 at most.

When you have more variables than the allowable number, you can try data reduction such as principle components or the propensity score, which we will learn later in the next chapter.

More advanced tool to account for over-fitting is "shrinkage" analysis. Problems due to over-fitting include inflation of both type I and type II error as stated on page 14 by Concato. Result of inflation type I errors, one may erroneously claim association when in fact there is no such an association. In this case, parameter estimate may be over-estimated (further away from the null value) and p-value tends to be smaller. Type II errors, one may claim there is no association when in fact there is. Parameter estimate tends to be smaller (closer to the null) and p-value tends to be bigger than actual. Inflation of type I error is more problematic than type II, because once the association is claimed, it is hard to disclaim such a finding.

Shrinkage analysis can numerically assess degree of over-fitting using bootstrap computation method, which quantify a degree of exaggeration made in parameter estimates in your analysis. For example when your data suggests that reduction in Hba1c is greater than 0.8% with intervention than control, true effect (the effect that other people are plausible to detect with similar dataset) was in fact 0.6%, shrinkage analysis quantify degree of over-fitting by 0.8-0.6/0.8=25%



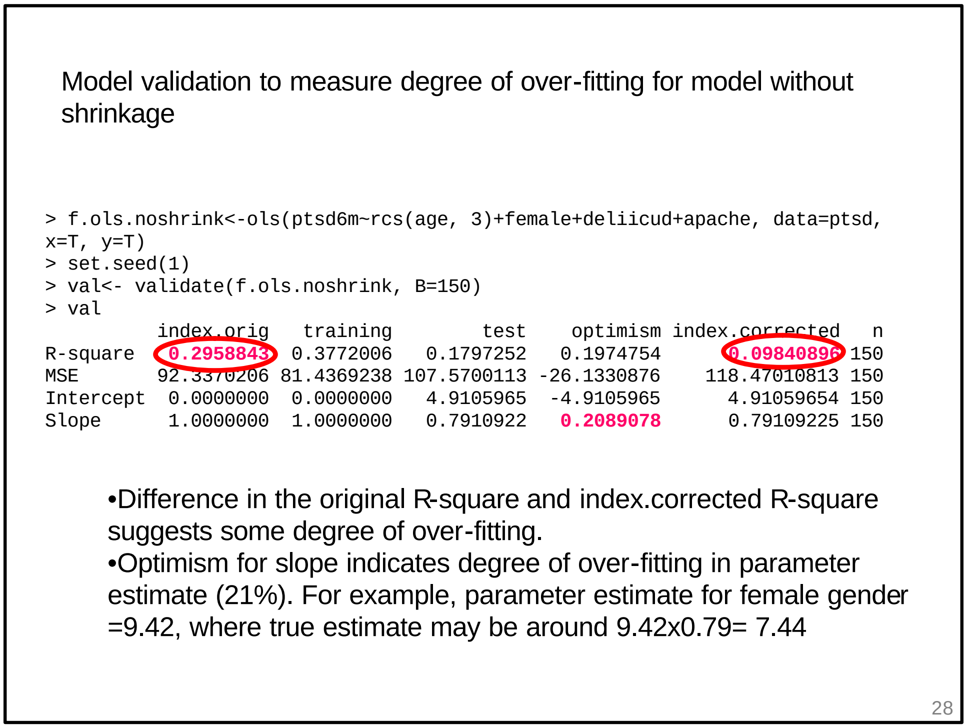
Pre-specified model without shrinkage

```
>f.ols.noshrink<-ols(ptsd6m~rcs(age, 3)+female+deliicud+apache, data=ptsd,
x=T, y=T)
>anova(f.ols.noshrink)
                Analysis of Variance          Response: ptsd6m

 Factor     d.f. Partial SS MS         F    P
 age         2    997.05235 498.52617 4.65 0.0158
  Nonlinear  1    551.71298 551.71298 5.14 0.0293
 female      1    740.68974 740.68974 6.90 0.0125
 deliicud    1    346.45687 346.45687 3.23 0.0805
 apache      1     39.99774  39.99774 0.37 0.5453
 REGRESSION  5   1668.48486 333.69697 3.11 0.0192
 ERROR      37   3970.49188 107.31059

>f.ols.noshrink
Coefficients:
            Value Std. Error       t Pr(>|t|)
Intercept  5.8744    14.0978  0.4167  0.67931
age        0.3352     0.2665  1.2576  0.21641
age'      -0.7688     0.3391 -2.2674  0.02930
female     9.4281     3.5886  2.6272  0.01245
deliicud   1.8806     1.0466  1.7968  0.08053
apache    -0.1259     0.2062 -0.6105  0.54525
```

27

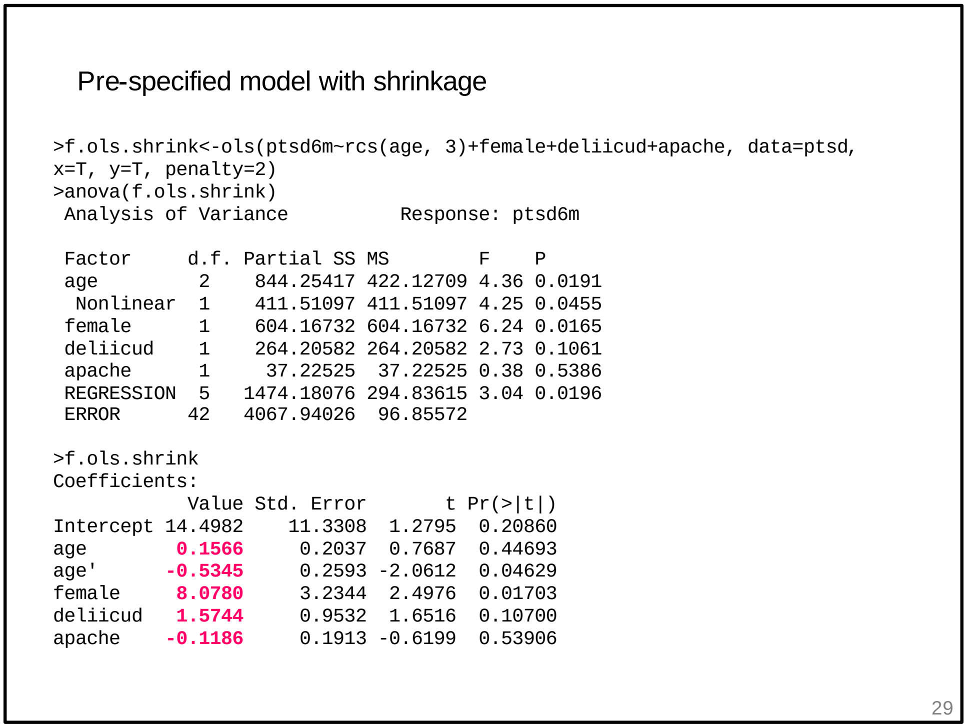


Model validation to measure degree of over-fitting for model without shrinkage

```
> f.ols.noshrink<-ols(ptsd6m~rcs(age, 3)+female+deliicud+apache, data=ptsd,
x=T, y=T)
> set.seed(1)
> val<- validate(f.ols.noshrink, B=150)
> val
          index.orig   training        test    optimism index.corrected   n
R-square   0.2958843  0.3772006   0.1797252   0.1974754      0.09840896 150
MSE       92.3370206 81.4369238 107.5700113 -26.1330876    118.47010813 150
Intercept  0.0000000  0.0000000   4.9105965  -4.9105965      4.91059654 150
Slope      1.0000000  1.0000000   0.7910922   0.2089078      0.79109225 150
```

- •Difference in the original R-square and index.corrected R-square suggests some degree of over-fitting.
- •Optimism for slope indicates degree of over-fitting in parameter estimate (21%). For example, parameter estimate for female gender =9.42, where true estimate may be around 9.42x0.79= 7.44

28



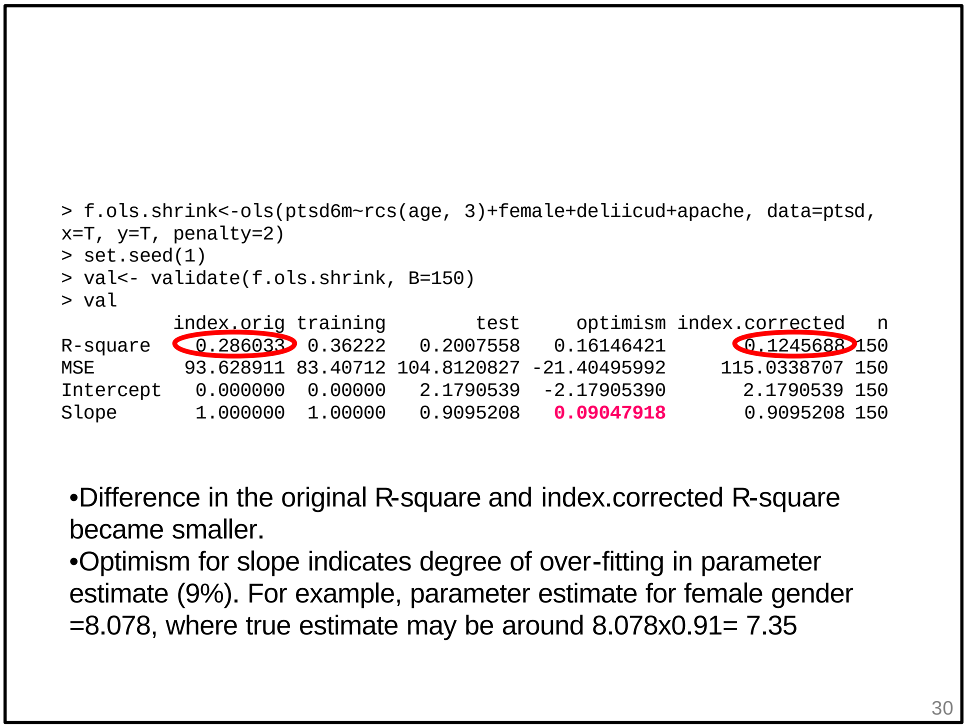
Pre-specified model with shrinkage

```
>f.ols.shrink<-ols(ptsd6m~rcs(age, 3)+female+deliicud+apache, data=ptsd,
x=T, y=T, penalty=2)
>anova(f.ols.shrink)
 Analysis of Variance          Response: ptsd6m

 Factor     d.f. Partial SS MS        F    P
 age         2    844.25417 422.12709 4.36 0.0191
  Nonlinear  1    411.51097 411.51097 4.25 0.0455
 female      1    604.16732 604.16732 6.24 0.0165
 deliicud    1    264.20582 264.20582 2.73 0.1061
 apache      1     37.22525  37.22525 0.38 0.5386
 REGRESSION  5   1474.18076 294.83615 3.04 0.0196
 ERROR      42   4067.94026  96.85572

>f.ols.shrink
Coefficients:
            Value Std. Error       t Pr(>|t|)
Intercept 14.4982    11.3308  1.2795  0.20860
age        0.1566     0.2037  0.7687  0.44693
age'      -0.5345     0.2593 -2.0612  0.04629
female     8.0780     3.2344  2.4976  0.01703
deliicud   1.5744     0.9532  1.6516  0.10700
apache    -0.1186     0.1913 -0.6199  0.53906
```

29



```
> f.ols.shrink<-ols(ptsd6m~rcs(age, 3)+female+deliicud+apache, data=ptsd,
x=T, y=T, penalty=2)
> set.seed(1)
> val<- validate(f.ols.shrink, B=150)
> val
          index.orig training        test     optimism index.corrected   n
R-square    0.286033  0.36222   0.2007558   0.16146421       0.1245688 150
MSE        93.628911 83.40712 104.8120827 -21.40495992     115.0338707 150
Intercept   0.000000  0.00000   2.1790539  -2.17905390       2.1790539 150
Slope       1.000000  1.00000   0.9095208   0.09047918       0.9095208 150
```

•Difference in the original R-square and index.corrected R-square became smaller.
•Optimism for slope indicates degree of over-fitting in parameter estimate (9%). For example, parameter estimate for female gender =8.078, where true estimate may be around 8.078x0.91= 7.35

30