General aims of the STRATOS INITIATIVE

Georg Heinze[1] and Willi Sauerbrei[2]
for the STRATOS initiative

[1]CeMSIIS-Section for Clinical Biometrics,
Medical University of Vienna, Austria

[2]Institute for Medical Biometry and Medical Informatics,
Medical Center - University of Freiburg, Germany

Statistical methodology – Current situation

- Substantial development over last decades
- Computer facilities
- Assess properties of complex models using simulation studies
- Resampling and Bayesian methods now easily available
- Wealth of new statistical software packages

Unfortunately, many sensible improvements are ignored in routine analyses

Why are our improvements ignored?

Highly sophisticated statistical methods

Statisticians with high expertise (in one particular area)

Routine: Simple methods, ignore dangers of wrong methods

Analysts with good statistical education

Data analysts with little statistical training

Why are our improvements ignored?

Highly sophisticated statistical methods, methods, methods

Statisticians with high expertise (in one particular area)

Routine: Simple methods, ignore dangers of wrong methods

Analysts with good statistical education

Data analysts with little statistical training

improvements ignored?

Highly sophisticated statistical methods, methods, methods, methods, methods, …

Statisticians with
high expertise
(in one particular area)

Routine:
Simple methods,
ignore dangers
of wrong
methods

Analysts with good statistical education

Data analysts with little statistical training

Why are our improvements ignored?

Highly sophisticated statistical methods

St... ...ns with ...tise ...particular area)

Needed: GUIDANCE!

Analysts with good s... ...cation

Data analysts with little statistical training

Routine: Simple methods, ignore dangers of wrong methods

STRengthening Analytical Thinking for Observational Studies: the STRATOS initiative

Willi Sauerbrei,[a*†] Michal Abrahamowicz,[b]
Douglas G. Altman,[c] Saskia le Cessie,[d] and[‡] James Carpenter[e]
on behalf of the STRATOS initiative

Statistics in Medicine 2014

2011	ISCB Ottawa, Epidemiology Sub-Comm.	Preliminary ideas
2012	ISCB Bergen	Discussions, SG
2013	ISCB Munich	Initiative launched
2014-16	ISCB	Invited Sessions
2016	Banff	Workshop
2016	IBC Victoria	Invited Session
2016	HEC Munich	Invited Session
2017	IBS-EMR Thessaloniki	Invited Session
2017	CEN-ISBS Vienna	Invited Session

http://www.stratos-initiative.org/

Basic information

Topic Group		Chairs and further members	
1	**Missing data**	Chairs:	**James Carpenter, Kate Lee**
		Members:	Melanie Bell, Els Goetghebeur, Joe Hogan, Rod Little, Andrea Rotnitzky, Kate Tilling, Ian White
2	**Selection of variables and functional forms in multivariable analysis**	Chairs:	**Michal Abrahamowicz, Aris Perperoglou, Willi Sauerbrei**
		Members:	Heiko Becher, Harald Binder, Frank Harrell, Georg Heinze, Patrick Royston, Matthias Schmid
3	**Initial data analysis**	Chairs:	**Marianne Huebner, Saskia le Cessie, Werner Vach**
		Members:	Maria Blettner, Dianne Cook, Heike Hofmann, Hermann-Josef Huss, Lara Lusa
4	**Measurement error and misclassification**	Chairs:	**Laurence Freedman, Victor Kipnis**
		Members:	Raymond Carroll, Veronika Deffner, Kevin Dodd, Paul Gustafson, Ruth Keogh, Helmut Küchenhoff, Pamela Shaw, Janet Tooze
5	**Study design**	Chairs:	**Mitchell Gail**
		Members:	Doug Altman, Gary Collins, Luc Duchateau, Neil Pearce, Peggy Sekula, Elizabeth Williamson, Mark Woodward
6	**Evaluating diagnostic tests and prediction models**	Chairs:	**Gary Collins, Carl Moons, Ewout Steyerberg**
		Members:	Patrick Bossuyt, Petra Macaskill, Ben van Calster, Andrew Vickers
7	**Causal inference**	Chairs:	**Els Goetghebeur**
		Members:	Bianca De Stavola, Saskia le Cessie, Niels Keiding, Erica Moodie, Ingeborg Waernbaum, Michael Wallace
8	**Survival analysis**	Chairs:	**Michal Abrahamowicz, Per Kragh Andersen, Terry Therneau**
		Members:	Richard Cook, Pierre Joly, Torben Martinussen, Maja Pohar-Perme, Jeremy Taylor
9	**High-dimensional data**	Chairs:	**Lisa McShane, Joerg Rahnenfuehrer**
		Members:	Axel Benner, Harald Binder, Anne-Laure Boulesteix, Tomasz Burzykowski, W. Evan Johnson, Lara Lusa, Stefan Michiels, Sherri Rose

Cross-cutting panels

Panels		Chairs
1	Glossary (GP)	Simon Day, Marianne Huebner, Jim Slattery
2	Data Sets (DP)	Saskia Le Cessie, Aris Perperoglou, Hermann Huss
3	Publications (PP)	Stephen Walter Co- Chairs: Bianca De Stavola, Mitchell Gail, Petra Macaskill
4	New Membership (MP)	James Carpenter, Willi Sauerbrei
5	Website (WP)	Joerg Rahnenfuehrer, Willi Sauerbrei
6	Literature Review (RP)	Gary Collins, Carl Moons
7	Simulation Studies (SP)	Michal Abrahamowicz, Harald Binder
8	Contact with Other Societies and Organizations (OP)	Willi Sauerbrei
9	Knowledge Transfer (TP)	Suzanne Cadarette

Why many researchers misuse variable selection– *and how to prevent this*

Georg Heinze and Daniela Dunkler

for STRATOS Topic Group 2

Medical University of Vienna

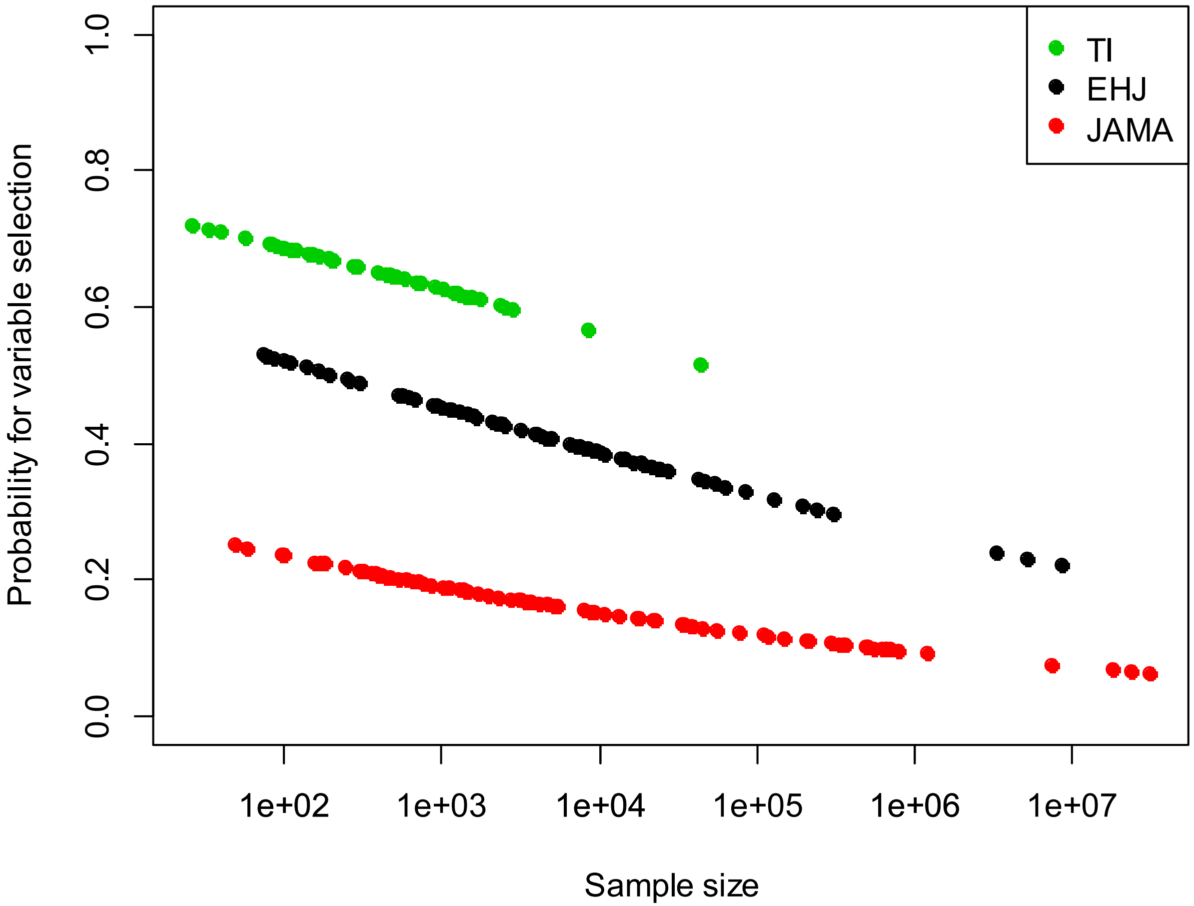
CeMSIIS – Section for Clinical Biometrics

Current practice of variable selection

Variable	JAMA Internal Medicine (IF=14.00)	European Heart Journal (IF=15.05)	Transplant International (IF=2.84)
A. Original articles 2015	137	132	89
B. Multivariable models	94	75	49
C. Variable selection (% of B)	17%	37%	65%
Univariate selection (% of B)	5%	21%	39%
Stepwise methods (% of B)	13%	23%	33%
Univariate filtering, then stepwise selection (% of B)	3%	8%	6%
Stability evaluation	0	0	0
Median sample size (in B)	4,396	4,319	295

Current practice of variable selection

- Modeling the probability for variable selection by journal and sample size:



The 5 myths about variable selection

1. The number of variables in a model should be reduced until there are 10 events per variable.
2. Only variables with proven univariable-model significance should be included in a multivariable model.
3. Non-significant effects should be eliminated from a model.
4. Selected-model p-values are valid.
5. Variable selection simplifies analysis.

➔ Probably because of these myths univariate selection is so popular.

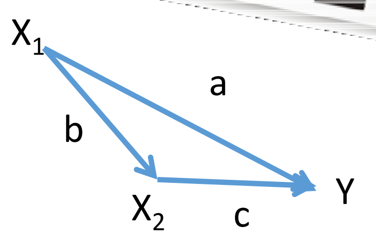
Myth 1: reduce until 10 events per variable

- Often a univariate ‚filter' is applied to reduce the variables that are included in a multivariable model
- But this ‚filter' is using the outcome data - > subject to sampling error
- Ignoring this uncertainty leads to problems
- Better: use only pre-existing knowledge to filter variables

Myth 2: include only univariately significant variables

- Easy. (You can do that with any software.)
- Retraceable.

- Problematic (see also Sun et al, JClinEpi 1996):
- The univariate effect of X_1 on Y is $a + bc$.



a	b	c	Consequence
Pos.	Pos.	Neg.	X_1 falsely not selected (if $a = -bc$)
0	Pos./Neg.	Pos./Neg.	X_1 falsely selected.
Pos./neg	0	Pos./neg	X_1 correctly selected (only if $b = 0$ or $c = 0$).

➔Univariate selection works only with uncorrelated variables.

Myth 3: remove non-significant variables

- It is commonly believed that ‚non-significant' variables must be removed as they add ‚noise' or even ‚bias' to the model

- ‚In multivariable analysis, only ABC1 and XYZ2 predicted the outcome.'

- Reverse argument: ‚X is not selected = X is not a predictor'

Background knowledge: simple illustrative simulations

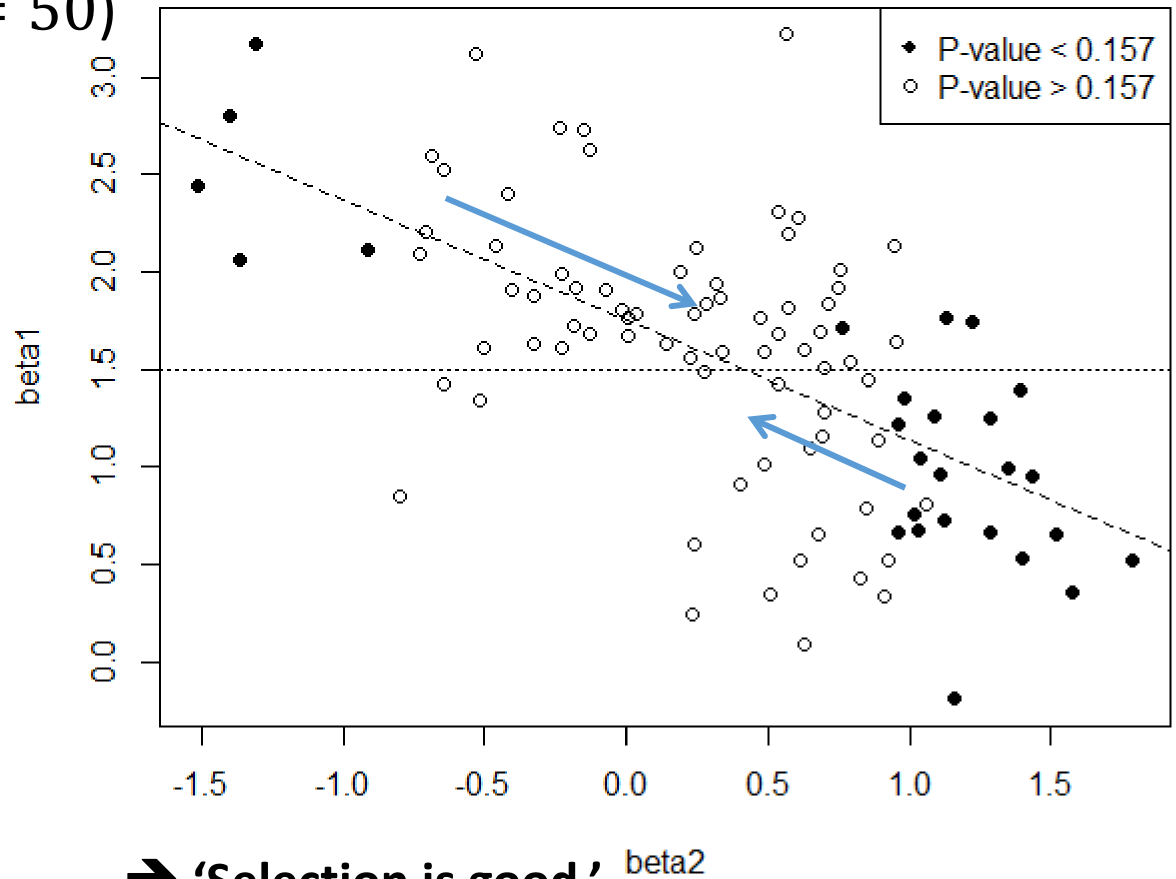
- Should X_2 be eliminated from the model? (simulation with $N = 50$)

True $\beta_1 = 1.5, \beta_2 = \mathbf{0.3}$

A weak β_2:

Setting it to 0 will more often push $\hat{\beta}_1$ towards its true value than away from it.
➔ Shrinkage effect on $\hat{\beta}_1$!

β_2:



➔ **'Selection is good.'**

Background knowledge: simple illustrative simulations

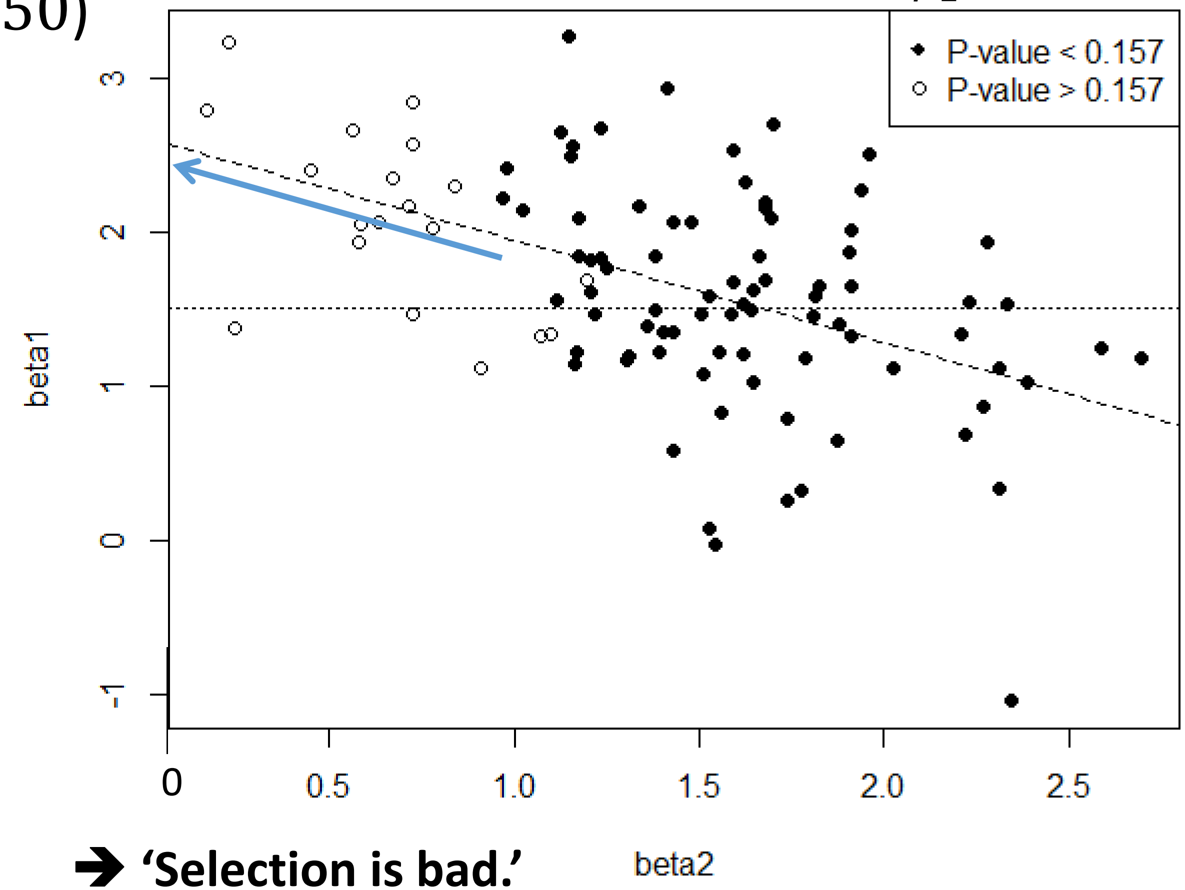
- Should X_2 be eliminated from the model? (simulation with $N = 50$)

True $\beta_1 = 1.5, \beta_2 = \mathbf{1.5}$

A strong β_2:

Setting it to 0 will always push $\hat{\beta}_1$ away from its true value.

β_2:



➔ **'Selection is bad.'**

Myth 4: Selected-model based p-values are valid

- After selection, software routinely reports model based p-values from the finally selected models
- These p-values are grossly misleading (biased low)

- Ignored:
  - uncertainty in selection decisions
  - multiplicity by performing several decisions step-by-step
  - At each step, p-value for β_j tests a different hypothesis!

- Better:
  - For inference, just use the p-values from the full model
  - (you considered all those variables for adjustment!)

Myth 5: Variable selection simplifies it

- Simple model – complex model

- But: additional uncertainty is introduced

- This additional uncertainty should be quantified (Heinze et al, 2017):
  - Selection probabilities of variables
  - Selection probabilities of models
  - Bias conditional on selection
  - RMSD ratios
  - Median coefficient, percentile confidence intervals
- The bootstrap (Sauerbrei and Schumacher, 1992) or subsampling (De Bin et al, 2015) can be used for this

The 5 myths: and what should change

1. The number of variables in a model should be reduced until there are 10 events per variable.
   Resp: No, there should be >>10 events per candidate variable.
2. Only variables with proven univariable-model significance should be included in a multivariable model.
   Resp: No, univariable-model significance can be strongly misleading as criterion for inclusion in a multivariable model.
3. Non-significant effects should be eliminated from a model.
   Resp: No, non-significant effects do not harm a model.
4. Selected-model based p-values are valid.
   Resp: No, P-values after model selection are almost impossible to estimate.
5. Variable selection simplifies analysis.
   Resp: No, stability investigations are needed and must become part of routine software output.

An example

Table 4 Body fat study: full model, model selected by backward elimination with a significance level of 0.157 (AIC selection), and some bootstrap-derived quantities useful for assessing model uncertainty.

Predictors	Full model		Bootstrap inclusion frequency (%)	Selected model		RMSD ratio	Relative conditional bias (%)	Bootstrap median	Bootstrap 2.5th percentile	Bootstrap 97.5th percentile
	Estimate	Standard error		Estimate	Standard error					
(Intercept)	4.14	23.27	100.0	5.95	8.15	1.06		4.27	-48.49	50.40
abdomen	0.90	0.09	100.0	0.87	0.06	1.06	-1.0	0.89	0.69	1.06
wrist	-1.84	0.53	97.5	-1.73	0.48	1.08	-1.5	-1.81	-2.79	-0.61
age	0.07	0.03	84.6	0.06	0.02	1.14	+5.2	0.07	0.00	0.13
height	-0.11	0.07	68.4	-0.13	0.05	1.14	+37.4	-0.11	-0.25	0.00
neck	-0.40	0.23	62.4	-0.33	0.22	1.24	+29.8	-0.38	-0.81	0.00
forearm	0.28	0.21	55.3	0.36	0.19	1.13	+46.4	0.28	0.00	0.64
thigh	0.17	0.15	49.7			1.14	+67.0	0.00	0.00	0.48
chest	-0.13	0.11	49.4	-0.14	0.09	1.14	+66.0	0.00	-0.34	0.00
biceps	0.17	0.17	43.8			1.15	+100.9	0.00	0.00	0.54
hip	-0.15	0.14	40.7			1.09	+86.7	0.00	-0.43	0.00
ankle	0.18	0.22	34.2			1.11	+84.2	0.00	-0.37	0.60
weight	-0.03	0.15	32.9			1.02	+383.3	0.00	-0.36	0.30
knee	-0.04	0.24	18.8			0.81	+203.2	0.00	-0.51	0.43

RMSD, root mean squared difference.

Johnson, 1996

An example

Table 4 Body fat study: full model, model selected by backward elimination with a significance level of 0.157 (AIC selection), and some bootstrap-derived quantities useful for assessing model uncertainty.

Predictors	Full model Estimate	Full model Standard error	Bootstrap inclusion frequency (%)	Selected model Estimate	Selected model Standard error	RMSD ratio	Relative conditional bias (%)	Bootstrap median	Bootstrap 2.5th percentile	Bootstrap 97.5th percentile
(Intercept)	4.14	23.27	100.0	5.95	8.15	1.06		4.27	-48.49	50.40
abdomen	0.90	0.09	100.0	0.87	0.06	1.06	-1.0	0.89	0.69	1.06
wrist	-1.84	0.53	97.5	-1.73	0.48	1.08	-1.5	-1.81	-2.79	-0.61
age	0.07	0.03	84.6	0.06	0.02	1.14	+5.2	0.07	0.00	0.13
height	-0.11	0.07	68.4	-0.13	0.05	1.14	+37.4	-0.11	-0.25	0.00
neck	-0.40	0.23	62.4	-0.33	0.22	1.24	+29.8	-0.38	-0.81	0.00
forearm	0.28	0.21	55.3	0.36	0.19	1.13	+46.4	0.28	0.00	0.64
thigh	0.17	0.15	49.7			1.14	+67.0	0.00	0.00	0.48
chest	-0.13	0.11	49.4	-0.14	0.09	1.14	+66.0	0.00	-0.34	0.00
biceps	0.17	0.17	43.8			1.15	+100.9	0.00	0.00	0.54
hip								0.00	-0.43	0.00
ankle								0.00	-0.37	0.60
weight								0.00	-0.36	0.30
knee								0.00	-0.51	0.43

RMSD, root me

Dear software developers, please implement this:
This will help to make researchers alert to the problems of variable selection.
Yours, Georg and Daniela

Johnson, 1996

References

- **Full tutorial 'Variable selection for statistical models: a review and recommendations for the practicing statistician'** with additional references: http://tinyurl.com/variable-selection-talk

- Heinze G, Wallisch C, Dunkler D. **Variable selection – a review and recommendations for the practicing statistician**. *Biometrical Journal*, invited review, submitted.
- Heinze G, Dunkler D. **Five myths about variable selection**. *Transplant International* 2017;30:6-10.

- De Bin R, Janitza S, Sauerbrei W, Boulesteix A. Subsampling versus bootstrapping in resampling-based model selection for multivariable regression. Biometrics 2016;72:272-80
- Johnson RW. Fitting percentage of body fat to simple body measurements. *Journal of Statistics Education* 1996; **4**. http://www.amstat.org/publications/jse/v4n1/datasets.johnson.html
- Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: Application to the Cox regression model. *Statistics in Medicine* 1992; **11**: 2093-2109
- Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *Journal of Clinical Epidemiology* 1996; **49**: 907-916.


CEN ISBS
CEN ISBS
Vienna 2017
Joint Conference on Biometrics &
Biopharmaceutical Statistics
28 August – 1 September 2017 in Vienna, Austria
Registration open!
Central European Network of the International Biometric Society
+ Austro-Swiss Region (ROeS)
+ German Region (IBS-DR)
+ Polish Region
International Society for Biopharmaceutical Statistics
©2016 CEN ISBS
Contact