





	Downloaded from http://ard.bmj.com/ on February 23, 2015	- Published by group.bmj.com			
		Viewpoint			
	Statistical review: frequently given comments				
	Stian Lydersen				
Handling editor Tore K Kvien	ABSTRACT	regression model, for example, 17 covariates in a	1		
Correspondence to Professor Stan Lydersen, Regional Centre for Child and Youth Mental Health and Child Welfare, Norwegian University of Science and Technology, Olax Kyres gate 9, P.O. Box 890, MTFS, Toordheim N-7491, Norway; stian.lydersen@ntru.no	From 2006 to 2014, 1 have carried out approximately 200 statistical reviews of manuscripts for <i>ARD</i> . My most frequent review comments concern the following: 1. Report how missing data were handled. 2. Limit the number of covariates in regression analyses. 3. Do not use stepwise selection of covariates. 4. Use analysis of covariance (ANCOVA) to adjust for baseline values in randomised controlled trials.	study with 64 cases. Traditional rules of thumb state that the ratio of cases per covariate ought to be in the size of order 10. Some authors recom- mend 15, some 20, others state that 5 is sufficient. In logistic regression and Cox regression, 10 events per variable is usually sufficient ² and in many situa- tions 5 events per variable is sufficient. ³ Note that in logistic regression this is not the total number of observations, but the smallest of the two outcome			
Received 30 June 2014 Revised 20 August 2014 Accepted 13 September 2014 Published Online First 26 September 2014	 Do not use ANCOVA to adjust for baseline values in observational studies. Dichotomising a continuous variable: a bad idea. Student's t test is better than non-parametric tests. 	groups. Similarly, in Cox regression, only the number of events excluding censored observations is counted as cases in this context.			
ov september 2014	 Do not use Yates' continuity correction. Mean (SD) is also relevant for non-normally distributed data. Report estimate, CI and (possibly) p value—in that order of importance. Post hoc power calculations—do not do it. Do not test for baseline imbalances in a randomised controlled trial. Report actual p values with 2 digits, maximum 3 	3. Do not use stepwise selection of covariates Automated variable selection procedures like step- wise selection used to be very popular. Today an increasing number of analysts criticise such methods. For example, ⁴ page 419 states: "There are several systematic, mechanical, and traditional algorithms for finding models (such as stepwise and best-subset regression) that lack logical and statis- tical iustification and that perform poorly in theory.			
	decimals. 14. Format for reporting Cls.	simulations and case studies One serious problem is that the P-values and standard errors will be downwardly biased, usually to a large degree". Selection of covariates should be based on the			
	mately 200 statistical reviews of manuscripts for	edge such as what is biologically plausible. Chapter			
	ARD. Some errors and weaknesses occur more	10 'Predictor selection' in the book5 gives good			

1. Report how missing data were handle

- · Report the amount of missing data
- ... and how this was handled.
- · Commonly used methods for handling missing data:
 - Complete case analysis
 - Single imputation methods like EM (the expectation-maximation algorithm)
 - Multiple imputation
 - Full information maximum likelihood
 - Linear mixed model in longitudinal studies
- LOCF (Last observation carried forward) is not unbiased under any sensible assumptions, and should not be used.









Lydersen, S. 2019. N Den norske legefore Lydersen, S. 2019. N Den norske legefore	langlende data - sje ning, 219, (3) 269 langlende uttrykk fi ning, 219, (3) 278	lden helt tilfeldig. Tidsskrift for or manglende data. Tidsskrift for	•
Engelsk term	Norsk term	Beskrivelse	
Missing completely at random (MCAR)	Mangler helt tilfeldig	Sannsynligheten for manglende data avhenger verken av observerte eller uobserverte data	
	Mangler betinget	Sannsynligheten for manglende data	7
Missing at random (MAR)	tilfeldig	avnenger bare av observerte data	



- For example, do not include17 covariates with 64 cases.
- The number of cases per covariate ought to be in the size of order 10. Some authors recommend15, some 20, others state that 5 is sufficient.
- In logistic regression, count cases in the smallest outcome group
- In Cox regression, count the number of events





(Katz 2006) In Preface: "Writing a second edition has given me the privilege of updating my thinking on multivariable analysis. The biggest change from the prior edition is that I have gone from being an "agnostic" on the topic of automatic variable selection algorithms (e.g. forward stepwise selection) to being against using them for explanatory models"

Katz, M.H. 2006. Multivariable analysis a practical guide for clinicians, 2nd ed. Cambridge, Cambridge University Press.

Tibshirani, R. 1996. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society.Series B, 58, (1) 267-288









- Dichotomising implies loss of information and reduced statistical power
- Dichotomising a covariate implies that the effect is a step-function changing only at the threshold.
- It can be sensible to dichotomise according to an established value such as a predefined clinical threshold
- Data-driven categorisation such as above/below the median is never a good idea.











10. Report estimate, CI and (possibly) p- valuein that order of importance

"When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid relying solely on statistical hypothesis testing, such as p-values, which fail to convey important information about effect size and precision of estimates".

"The Vancouver guidelines" www.icmje.org

















