XI. OTHER TOPICS

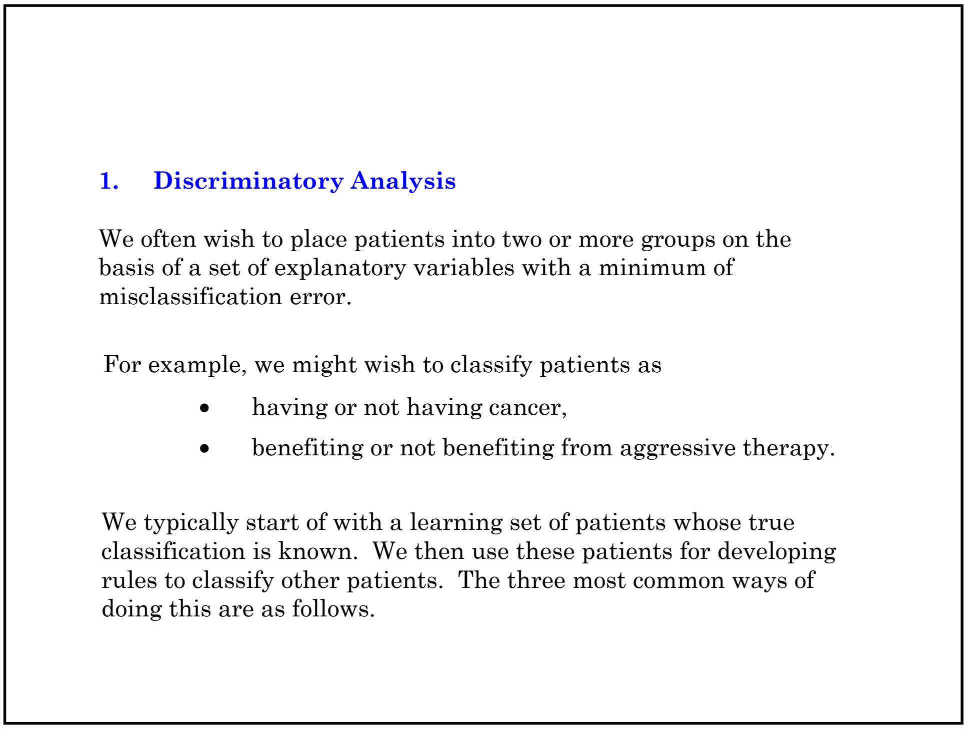
Complicated Statistics with Nasty Properties

Bootstrap Analysis

- Treat the sample as if it were the target population
- Sample repeatedly without replacement to obtain many samples of the same size as the real sample
- Calculate the test statistic for each sample
- Examine the variation of the test statistic among bootstrapped samples to assess its dispersion.

Multiple imputation of missing values

- Most statistical packages, including Stata do complete case analyses. That is they discard the data on any patient who is missing any model covariate.
- Multiple imputation is a method that adjusts for missing data by predicting missing values from non-missing covariates.
- Lead to unbiased results if the probability of the outcome of interest is not affected by whether a specific covariate is missing.
- Stata has a very comprehensive package for doing multiple imputation
- Particularly useful to adjust for missing values in confounding variables.



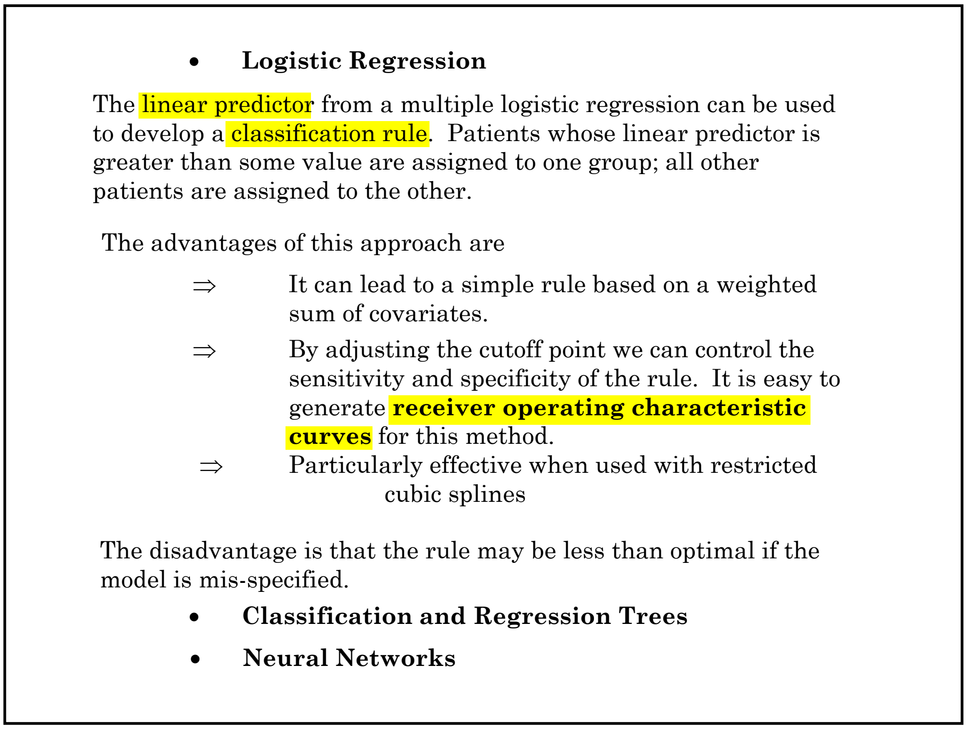
1. Discriminatory Analysis

We often wish to place patients into two or more groups on the basis of a set of explanatory variables with a minimum of misclassification error.

For example, we might wish to classify patients as

- having or not having cancer,
- benefiting or not benefiting from aggressive therapy.

We typically start of with a learning set of patients whose true classification is known. We then use these patients for developing rules to classify other patients. The three most common ways of doing this are as follows.



- **Logistic Regression**

The linear predictor from a multiple logistic regression can be used to develop a classification rule. Patients whose linear predictor is greater than some value are assigned to one group; all other patients are assigned to the other.

The advantages of this approach are

⇒ It can lead to a simple rule based on a weighted sum of covariates.

⇒ By adjusting the cutoff point we can control the sensitivity and specificity of the rule. It is easy to generate **receiver operating characteristic curves** for this method.

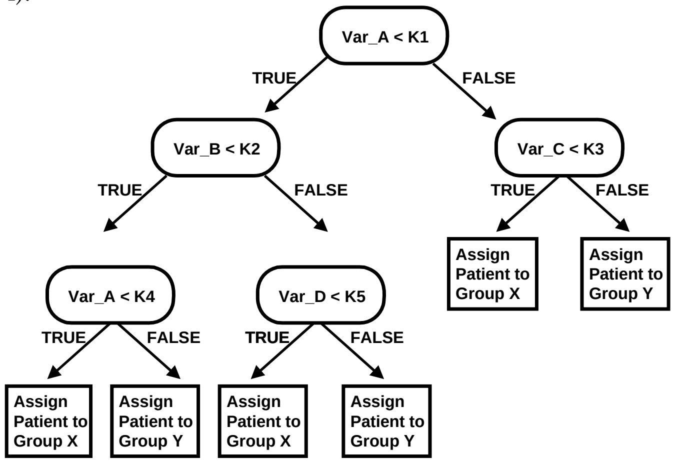
⇒ Particularly effective when used with restricted cubic splines

The disadvantage is that the rule may be less than optimal if the model is mis-specified.

- **Classification and Regression Trees**
- **Neural Networks**

2. Classification and Regression Trees (CART)

The basic idea here is to derive a tree that consists of a series of binary decisions that lead to patient classification (Breiman et al. 1984).

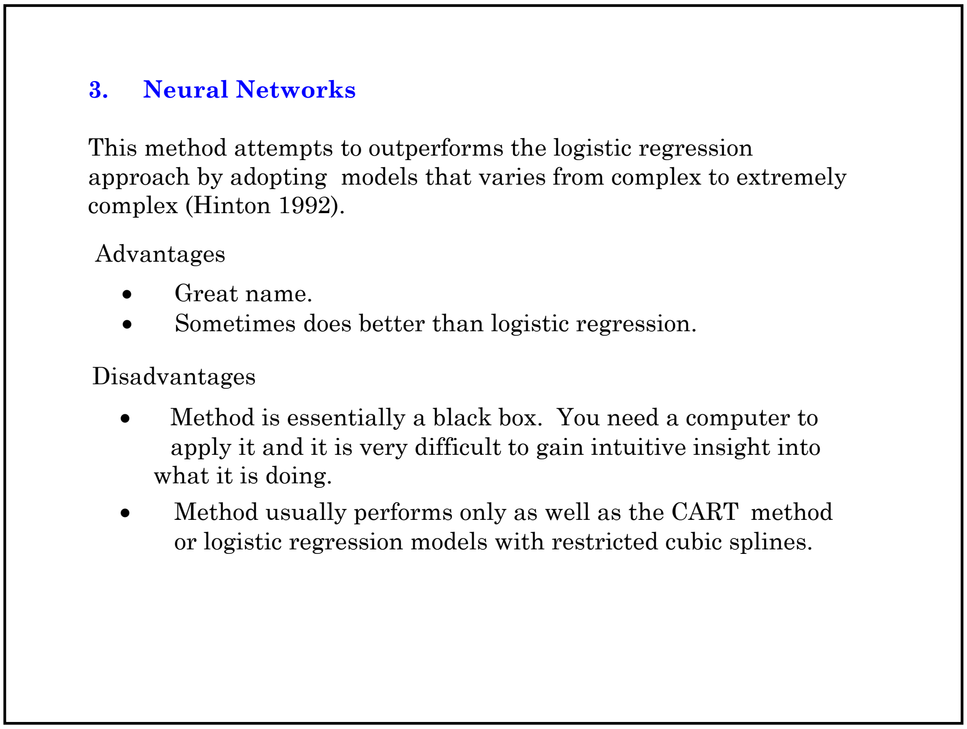

Var_A < K1
TRUE
FALSE
Var_B < K2
Var_C < K3
TRUE
FALSE
TRUE
FALSE
Assign Patient to Group X
Assign Patient to Group Y
Var_A < K4
Var_D < K5
TRUE
FALSE
TRUE
FALSE
Assign Patient to Group X
Assign Patient to Group Y
Assign Patient to Group X
Assign Patient to Group Y

The CART graphic indicates the degree of increased homogeneity induced by each split. Trees can then be pruned back to produce a classification rule that makes clinical sense and is fairly easy to remember.

The advantages of this method are

- It often does better than logistic regression when the model for the latter is poorly specified.
- It gives a rule that is intelligible to clinicians and can be judged by its clinical criteria.

A disadvantage is that, when applied to continuous covariates it looses information due to the fact that it dichotomizes the selected variable at each split.



3. Neural Networks

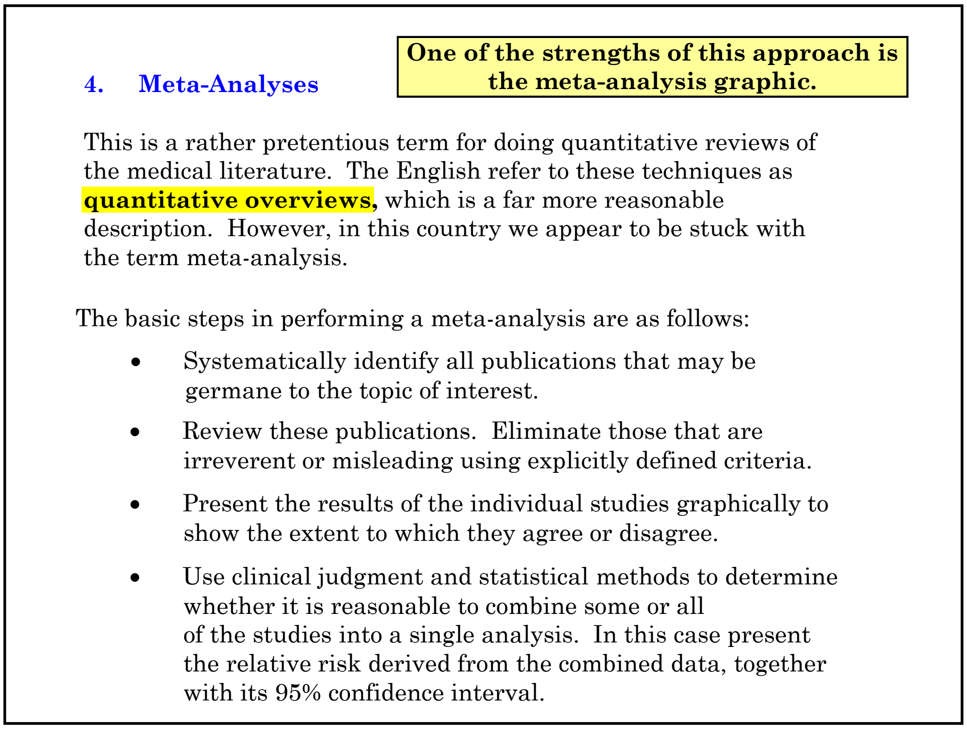
This method attempts to outperforms the logistic regression approach by adopting models that varies from complex to extremely complex (Hinton 1992).

Advantages

- Great name.
- Sometimes does better than logistic regression.

Disadvantages

- Method is essentially a black box. You need a computer to apply it and it is very difficult to gain intuitive insight into what it is doing.
- Method usually performs only as well as the CART method or logistic regression models with restricted cubic splines.



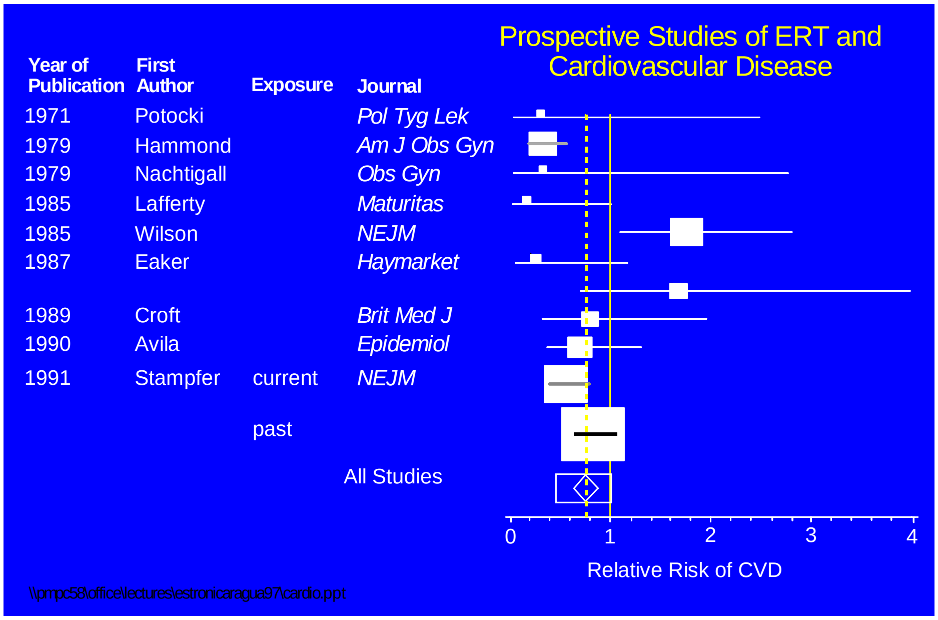
4. Meta-Analyses

One of the strengths of this approach is the meta-analysis graphic.

This is a rather pretentious term for doing quantitative reviews of the medical literature. The English refer to these techniques as **quantitative overviews,** which is a far more reasonable description. However, in this country we appear to be stuck with the term meta-analysis.

The basic steps in performing a meta-analysis are as follows:

- Systematically identify all publications that may be germane to the topic of interest.
- Review these publications. Eliminate those that are irreverent or misleading using explicitly defined criteria.
- Present the results of the individual studies graphically to show the extent to which they agree or disagree.
- Use clinical judgment and statistical methods to determine whether it is reasonable to combine some or all of the studies into a single analysis. In this case present the relative risk derived from the combined data, together with its 95% confidence interval.



Prospective Studies of ERT and Cardiovascular Disease

Year of Publication	First Author	Exposure	Journal
1971	Potocki		*Pol Tyg Lek*
1979	Hammond		*Am J Obs Gyn*
1979	Nachtigall		*Obs Gyn*
1985	Lafferty		*Maturitas*
1985	Wilson		*NEJM*
1987	Eaker		*Haymarket*
1989	Croft		*Brit Med J*
1990	Avila		*Epidemiol*
1991	Stampfer	current	*NEJM*
		past	
			All Studies

\\pmpc58\office\lectures\estronicaragua97\cardio.ppt

- In these graphs the relative risk from each study is displayed on a single line.
- Each relative risk or odds ratio is plotted as a square.
- The size of this square is proportional to the reciprocal of the variance of the log relative risk (often referred to as the study **information**).
- The 95% confidence interval for each study is depicted as a horizontal line.
- A vertical line depicts a weighted geometric mean of the studies. This mean is weighted by the information content of each study.
- One, or preferably two, 95% confidence intervals are drawn for this combined geometric mean. These confidence intervals are usually drawn as diamonds or squares. They are calculated using either a fixed effects or random effects model.

a) Fixed effects model for meta-analysis

This approach assumes that all studies are measuring the same risk in a comparable way, and that the only variation between studies is due to chance.

If this assumption is false it will overestimate the precision of the combined estimate.

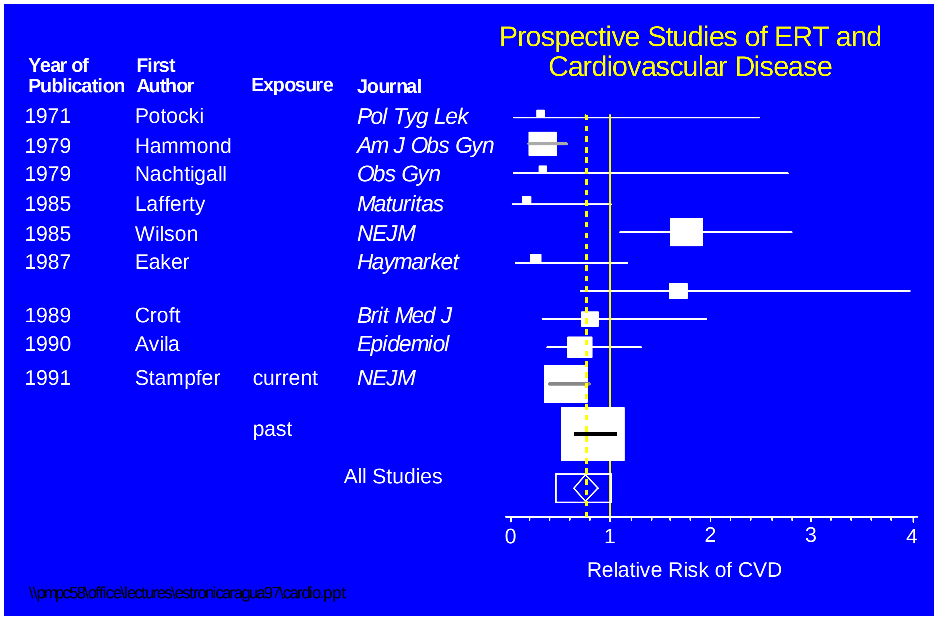
b) Random effects model for meta-analysis

This model assumes that that each study is estimating a different unknown relative risk that is specific to that study. These risks differ from one study to the next due to differences in study populations, study designs, or biases of one kind or another.

It assumes that these study-specific relative risks follow a log-normal distribution, and that the variation in the estimated relative risks is due both to variation in the study specific risk as well as intra-study variation of study subjects.

DerSimonian and Laird (1986) devised a way to estimate the confidence interval for the combined relative risk for this model.

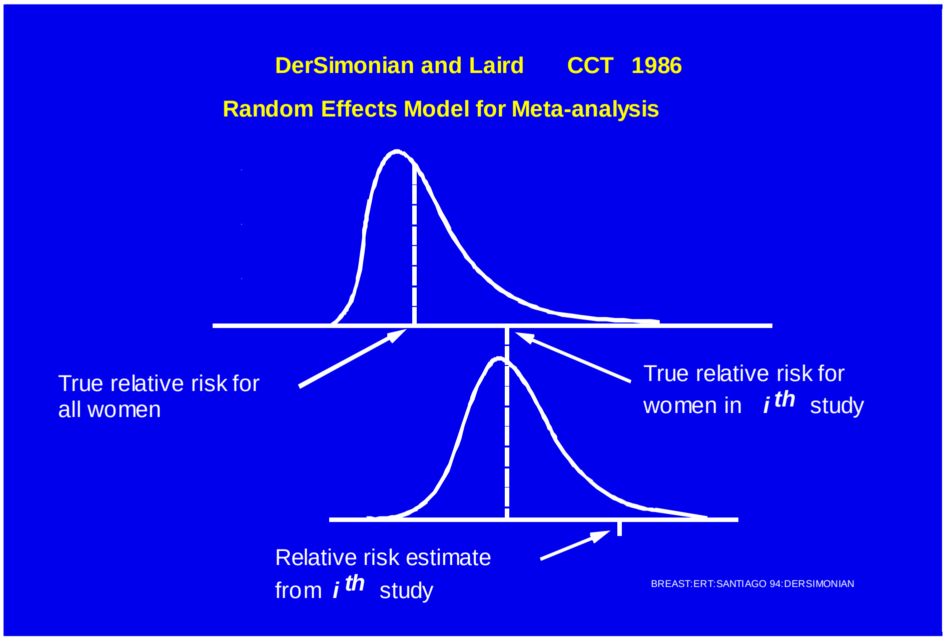
It is a good idea to plot both the fixed effects and random effects confidence intervals for the combined relative risk estimate. If these intervals disagree then the inter-study variation is greater than we would expect by chance and the studies are most likely estimating different risks. In this case we need to be very cautious about combining the results of these studies.



Prospective Studies of ERT and Cardiovascular Disease

Year of Publication	First Author	Exposure	Journal
1971	Potocki		*Pol Tyg Lek*
1979	Hammond		*Am J Obs Gyn*
1979	Nachtigall		*Obs Gyn*
1985	Lafferty		*Maturitas*
1985	Wilson		*NEJM*
1987	Eaker		*Haymarket*
1989	Croft		*Brit Med J*
1990	Avila		*Epidemiol*
1991	Stampfer	current	*NEJM*
		past	
			All Studies

\\pmpc58\office\lectures\estronicaragua97\cardio.ppt



DerSimonian and Laird CCT 1986

Random Effects Model for Meta-analysis

True relative risk for all women

True relative risk for women in *i* *th* study

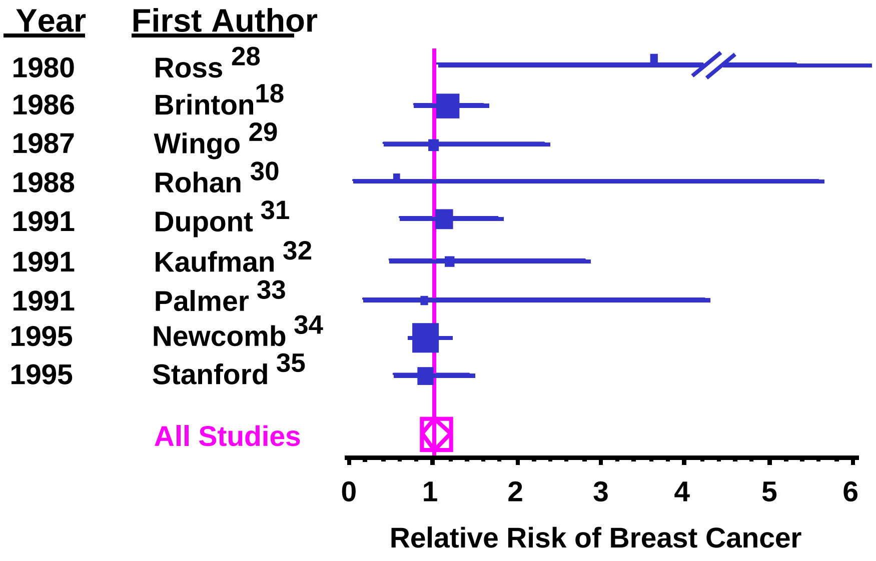
Relative risk estimate from *i* *th* study

BREAST:ERT:SANTIAGO 94:DERSIMONIAN

On the other hand, if these estimates agree then the studies are mutually consistent and there is no statistical reason not to combine them.

Women with a History of Benign Breast Disease

Breast cancer risk among ERT users compared to non-users



5. Publication bias

One of the ways that meta-analyses can be misleading is through publication bias. That is, papers may be more likely to be published if they show that a risk factor either increases or reduces some risk than if they find a relative risk near one.

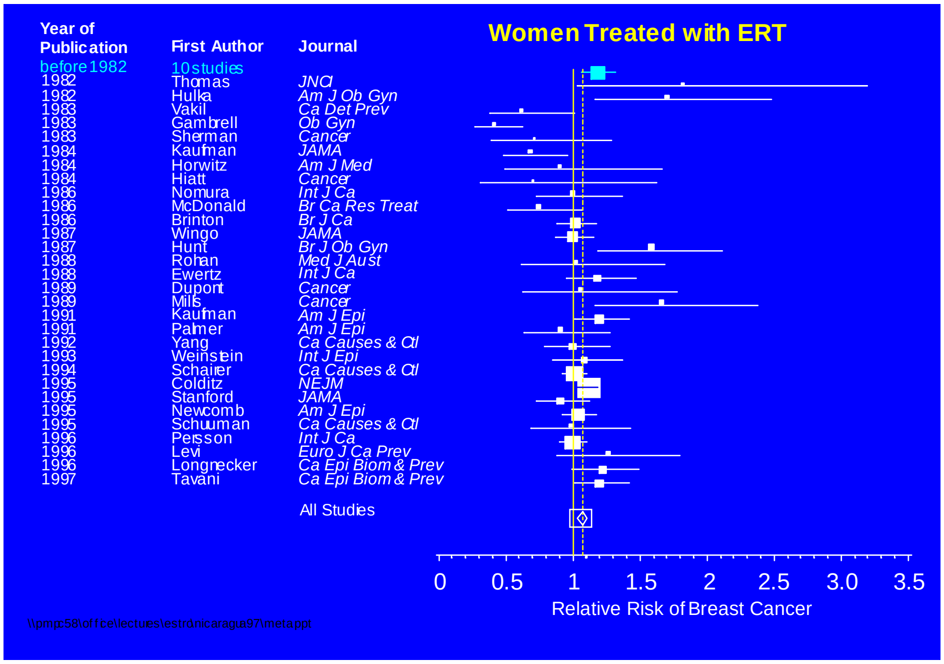
Small studies are more likely to be affected by publication bias than large ones.

6. Funnel graphs

One way to check for publication bias is to plot funnel graphs (Light & Pillemer 1984).

In these graphs we plot the standard error of the log relative risk against log relative risk. If this plot has a funnel shape we have evidence of publication bias

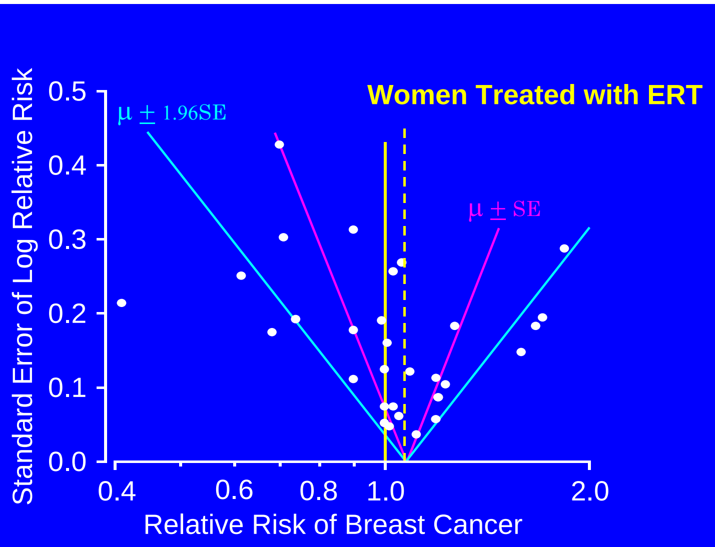
When this happens it may make sense to exclude studies with a standard error of the log relative risk that is greater than some value.



Women Treated with ERT

Year of Publication	First Author	Journal
before 1982	10 studies	
1982	Thomas	*JNCI*
1982	Hulka	*Am J Ob Gyn*
1983	Vakil	*Ca Det Prev*
1983	Gambrell	*Ob Gyn*
1983	Sherman	*Cancer*
1984	Kaufman	*JAMA*
1984	Horwitz	*Am J Med*
1984	Hiatt	*Cancer*
1986	Nomura	*Int J Ca*
1986	McDonald	*Br Ca Res Treat*
1986	Brinton	*Br J Ca*
1987	Wingo	*JAMA*
1987	Hunt	*Br J Ob Gyn*
1988	Rohan	*Med J Aust*
1988	Ewertz	*Int J Ca*
1989	Dupont	*Cancer*
1989	Mills	*Cancer*
1991	Kaufman	*Am J Epi*
1991	Palmer	*Am J Epi*
1992	Yang	*Ca Causes & Ctl*
1993	Weinstein	*Int J Epi*
1994	Schairer	*Ca Causes & Ctl*
1995	Colditz	*NEJM*
1995	Stanford	*JAMA*
1995	Newcomb	*Am J Epi*
1995	Schuuman	*Ca Causes & Ctl*
1996	Persson	*Int J Ca*
1996	Levi	*Euro J Ca Prev*
1996	Longnecker	*Ca Epi Biom & Prev*
1997	Tavani	*Ca Epi Biom & Prev*
		All Studies

\\pmpc58\office\lectures\estrd\nicaragua97\meta.ppt



\\pmpc58\office\lectures\estro\nicaragua97\funnel.ppt

Approaches to Extreme Multiple Comparisons Problems

- ❖ Permutation Tests
- ❖ Cross validation Methods
- ❖ False Discovery Rates
- ❖ Shrinkage Analysis
- ❖ Learning set Test Set Analyses

This course has been concerned with methods that are appropriate when the number of patients far exceeds the number of model parameters.

Diversity Among Statisticians

We all want to

Minimize probabilities of Type I errors

Minimize probabilities of Type II errors

All other things being equal, simple explanations are better than complex ones.

> *Science may be described as the art of systematic over-simplification — the art of discerning what we may with advantage omit.*
>
> Karl Popper

Today, reputable statisticians may disagree to some extent about the relative emphasis that should be placed on these three goals

XII. SUMMARY OF MULTIPLE REGRESSION METHODS

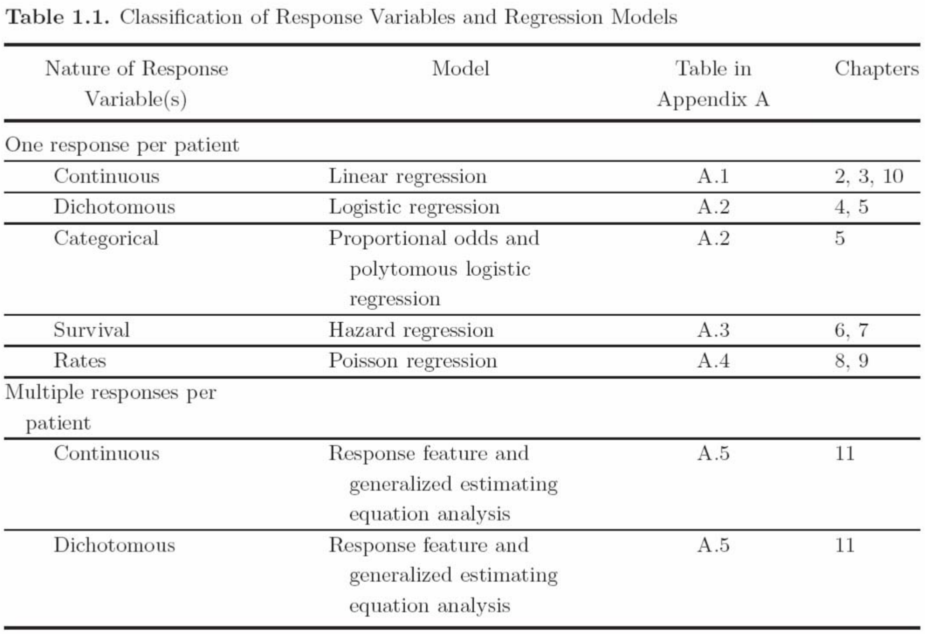


Table 1.1. Classification of Response Variables and Regression Models

Nature of Response Variable(s)	Model	Table in Appendix A	Chapters
One response per patient			
Continuous	Linear regression	A.1	2, 3, 10
Dichotomous	Logistic regression	A.2	4, 5
Categorical	Proportional odds and polytomous logistic regression	A.2	5
Survival	Hazard regression	A.3	6, 7
Rates	Poisson regression	A.4	8, 9
Multiple responses per patient			
Continuous	Response feature and generalized estimating equation analysis	A.5	11
Dichotomous	Response feature and generalized estimating equation analysis	A.5	11

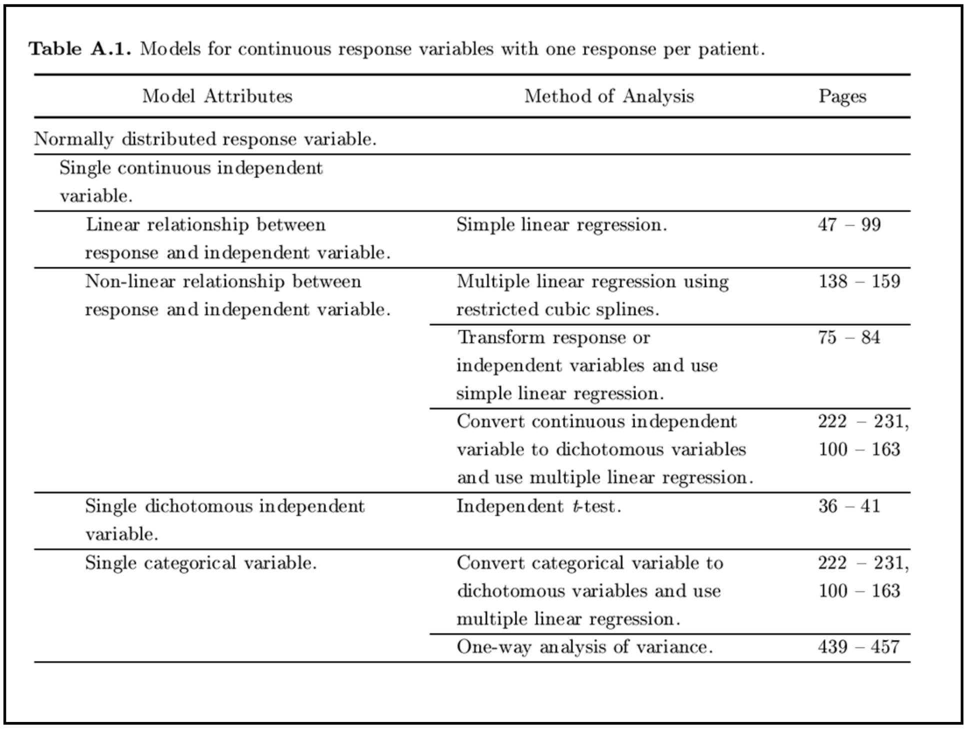


Table A.1. Models for continuous response variables with one response per patient.

Model Attributes	Method of Analysis	Pages
Normally distributed response variable.		
Single continuous independent variable.		
Linear relationship between response and independent variable.	Simple linear regression.	47 – 99
Non-linear relationship between response and independent variable.	Multiple linear regression using restricted cubic splines.	138 – 159
	Transform response or independent variables and use simple linear regression.	75 – 84
	Convert continuous independent variable to dichotomous variables and use multiple linear regression.	222 – 231, 100 – 163
Single dichotomous independent variable.	Independent *t*-test.	36 – 41
Single categorical variable.	Convert categorical variable to dichotomous variables and use multiple linear regression.	222 – 231, 100 – 163
	One-way analysis of variance.	439 – 457

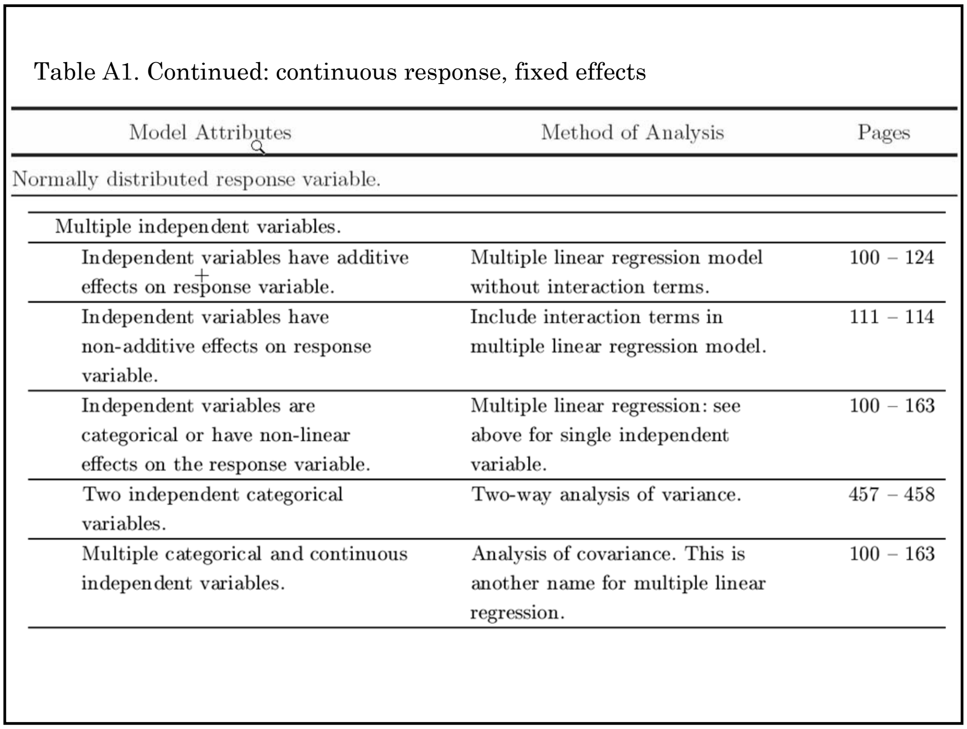


Table A1. Continued: continuous response, fixed effects

Model Attributes	Method of Analysis	Pages
Normally distributed response variable.		
Multiple independent variables.		
Independent variables have additive effects on response variable.	Multiple linear regression model without interaction terms.	100 – 124
Independent variables have non-additive effects on response variable.	Include interaction terms in multiple linear regression model.	111 – 114
Independent variables are categorical or have non-linear effects on the response variable.	Multiple linear regression: see above for single independent variable.	100 – 163
Two independent categorical variables.	Two-way analysis of variance.	457 – 458
Multiple categorical and continuous independent variables.	Analysis of covariance. This is another name for multiple linear regression.	100 – 163

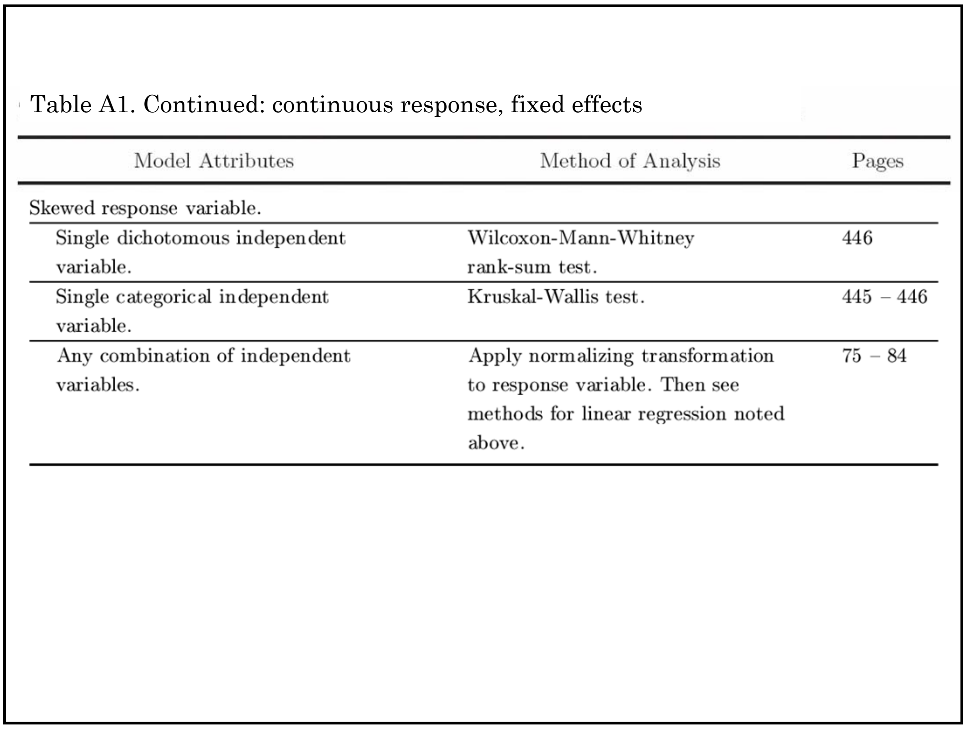


Table A1. Continued: continuous response, fixed effects

Model Attributes	Method of Analysis	Pages
Skewed response variable.		
Single dichotomous independent variable.	Wilcoxon-Mann-Whitney rank-sum test.	446
Single categorical independent variable.	Kruskal-Wallis test.	445 – 446
Any combination of independent variables.	Apply normalizing transformation to response variable. Then see methods for linear regression noted above.	75 – 84

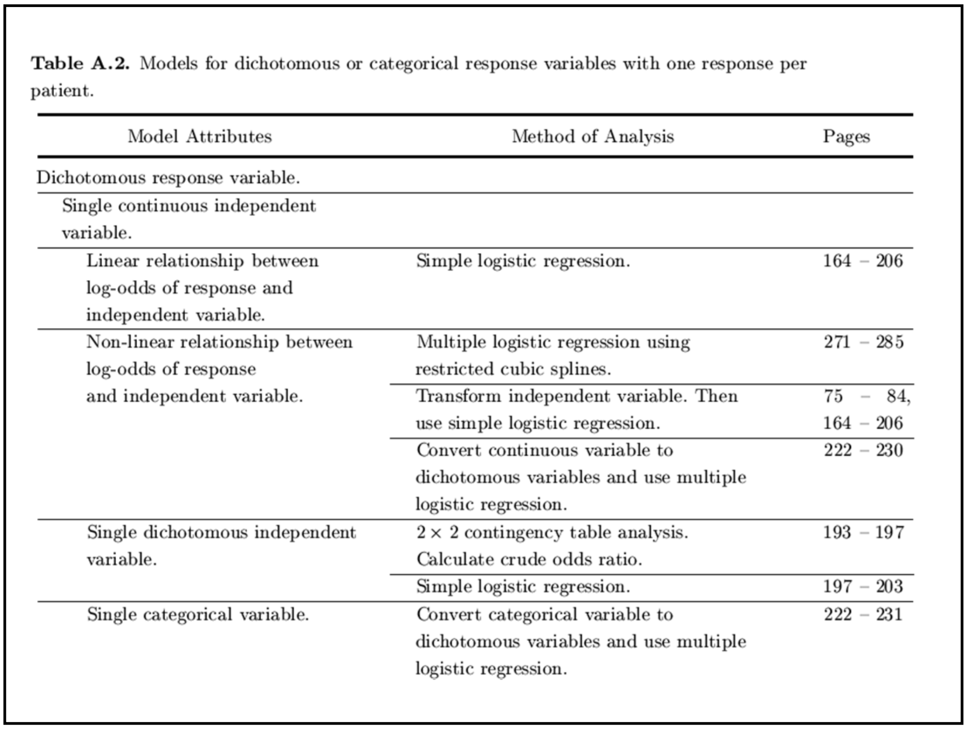
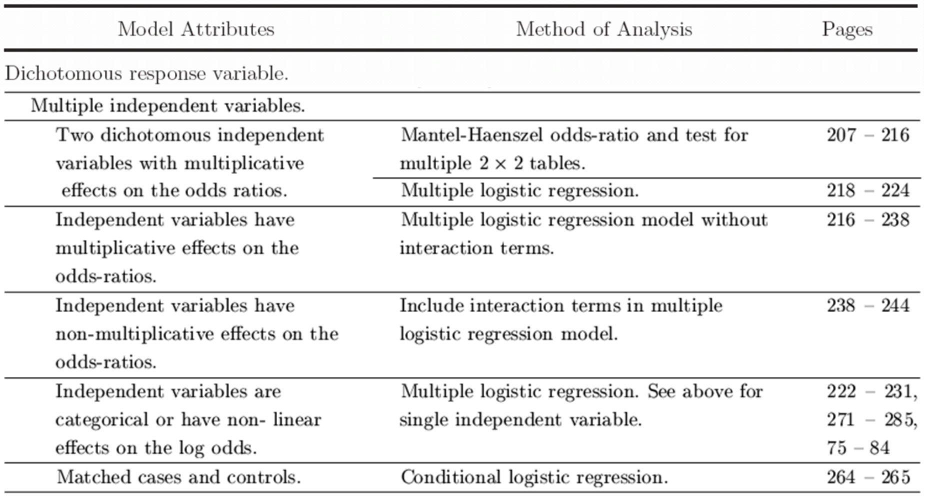


Table A.2. Models for dichotomous or categorical response variables with one response per patient.

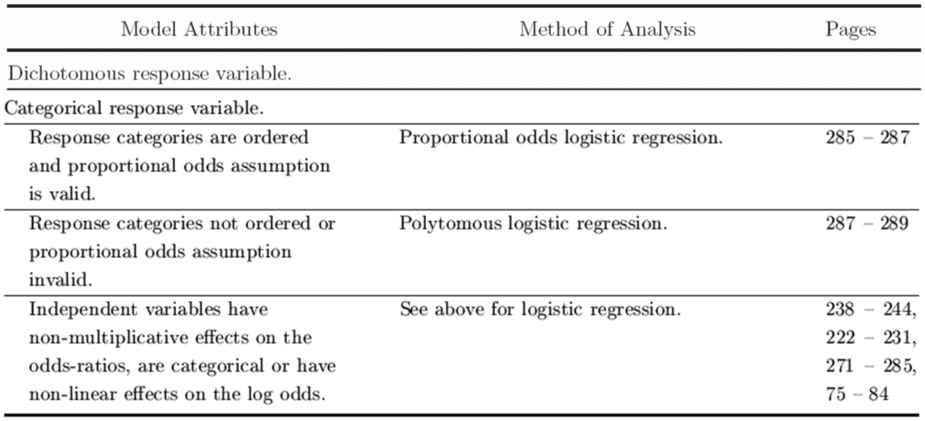
Model Attributes	Method of Analysis	Pages
Dichotomous response variable.		
Single continuous independent variable.		
Linear relationship between log-odds of response and independent variable.	Simple logistic regression.	164 – 206
Non-linear relationship between log-odds of response and independent variable.	Multiple logistic regression using restricted cubic splines.	271 – 285
	Transform independent variable. Then use simple logistic regression.	75 – 84, 164 – 206
	Convert continuous variable to dichotomous variables and use multiple logistic regression.	222 – 230
Single dichotomous independent variable.	2 × 2 contingency table analysis. Calculate crude odds ratio.	193 – 197
	Simple logistic regression.	197 – 203
Single categorical variable.	Convert categorical variable to dichotomous variables and use multiple logistic regression.	222 – 231

Table A2. Continued: dichotomous response, fixed effects



Model Attributes	Method of Analysis	Pages
Dichotomous response variable.		
Multiple independent variables.		
Two dichotomous independent variables with multiplicative effects on the odds ratios.	Mantel-Haenszel odds-ratio and test for multiple 2 × 2 tables.	207 – 216
	Multiple logistic regression.	218 – 224
Independent variables have multiplicative effects on the odds-ratios.	Multiple logistic regression model without interaction terms.	216 – 238
Independent variables have non-multiplicative effects on the odds-ratios.	Include interaction terms in multiple logistic regression model.	238 – 244
Independent variables are categorical or have non- linear effects on the log odds.	Multiple logistic regression. See above for single independent variable.	222 – 231, 271 – 285, 75 – 84
Matched cases and controls.	Conditional logistic regression.	264 – 265

Table A2. Continued: categorical response, fixed effects



Model Attributes	Method of Analysis	Pages
Dichotomous response variable.		
Categorical response variable.		
Response categories are ordered and proportional odds assumption is valid.	Proportional odds logistic regression.	285 – 287
Response categories not ordered or proportional odds assumption invalid.	Polytomous logistic regression.	287 – 289
Independent variables have non-multiplicative effects on the odds-ratios, are categorical or have non-linear effects on the log odds.	See above for logistic regression.	238 – 244, 222 – 231, 271 – 285, 75 – 84

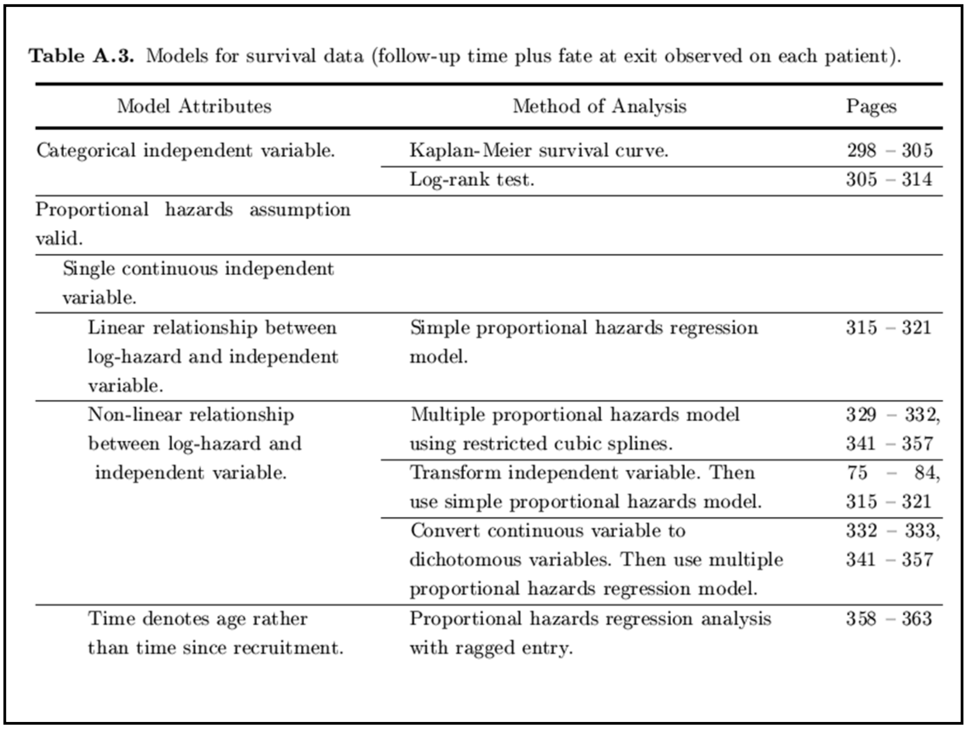


Table A.3. Models for survival data (follow-up time plus fate at exit observed on each patient).

Model Attributes	Method of Analysis	Pages
Categorical independent variable.	Kaplan-Meier survival curve.	298 – 305
	Log-rank test.	305 – 314
Proportional hazards assumption valid.		
Single continuous independent variable.		
Linear relationship between log-hazard and independent variable.	Simple proportional hazards regression model.	315 – 321
Non-linear relationship between log-hazard and independent variable.	Multiple proportional hazards model using restricted cubic splines.	329 – 332, 341 – 357
	Transform independent variable. Then use simple proportional hazards model.	75 – 84, 315 – 321
	Convert continuous variable to dichotomous variables. Then use multiple proportional hazards regression model.	332 – 333, 341 – 357
Time denotes age rather than time since recruitment.	Proportional hazards regression analysis with ragged entry.	358 – 363

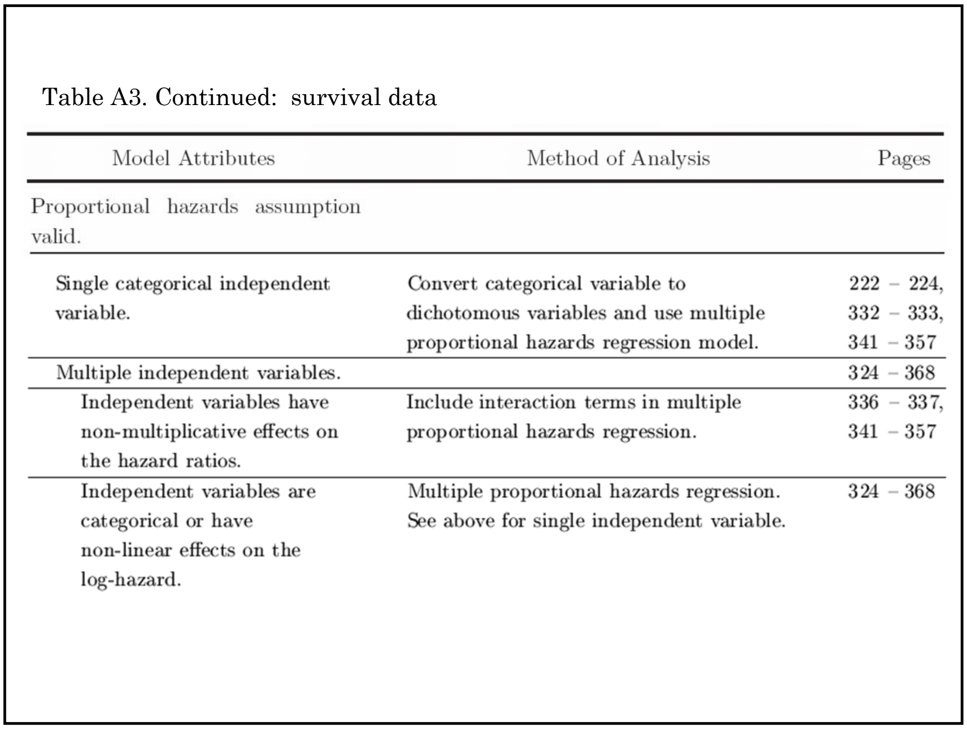


Table A3. Continued: survival data

Model Attributes	Method of Analysis	Pages
Proportional hazards assumption valid.		
Single categorical independent variable.	Convert categorical variable to dichotomous variables and use multiple proportional hazards regression model.	222 – 224, 332 – 333, 341 – 357
Multiple independent variables.		324 – 368
Independent variables have non-multiplicative effects on the hazard ratios.	Include interaction terms in multiple proportional hazards regression.	336 – 337, 341 – 357
Independent variables are categorical or have non-linear effects on the log-hazard.	Multiple proportional hazards regression. See above for single independent variable.	324 – 368

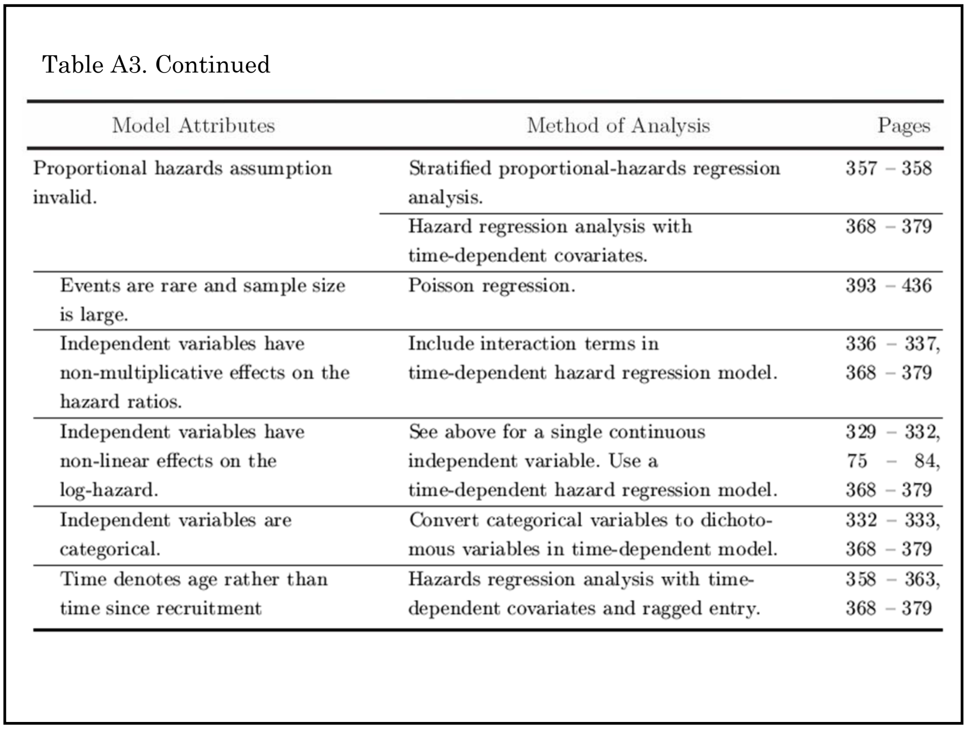


Table A3. Continued

Model Attributes	Method of Analysis	Pages
Proportional hazards assumption invalid.	Stratified proportional-hazards regression analysis.	357 – 358
	Hazard regression analysis with time-dependent covariates.	368 – 379
Events are rare and sample size is large.	Poisson regression.	393 – 436
Independent variables have non-multiplicative effects on the hazard ratios.	Include interaction terms in time-dependent hazard regression model.	336 – 337, 368 – 379
Independent variables have non-linear effects on the log-hazard.	See above for a single continuous independent variable. Use a time-dependent hazard regression model.	329 – 332, 75 – 84, 368 – 379
Independent variables are categorical.	Convert categorical variables to dichotomous variables in time-dependent model.	332 – 333, 368 – 379
Time denotes age rather than time since recruitment	Hazards regression analysis with time-dependent covariates and ragged entry.	358 – 363, 368 – 379

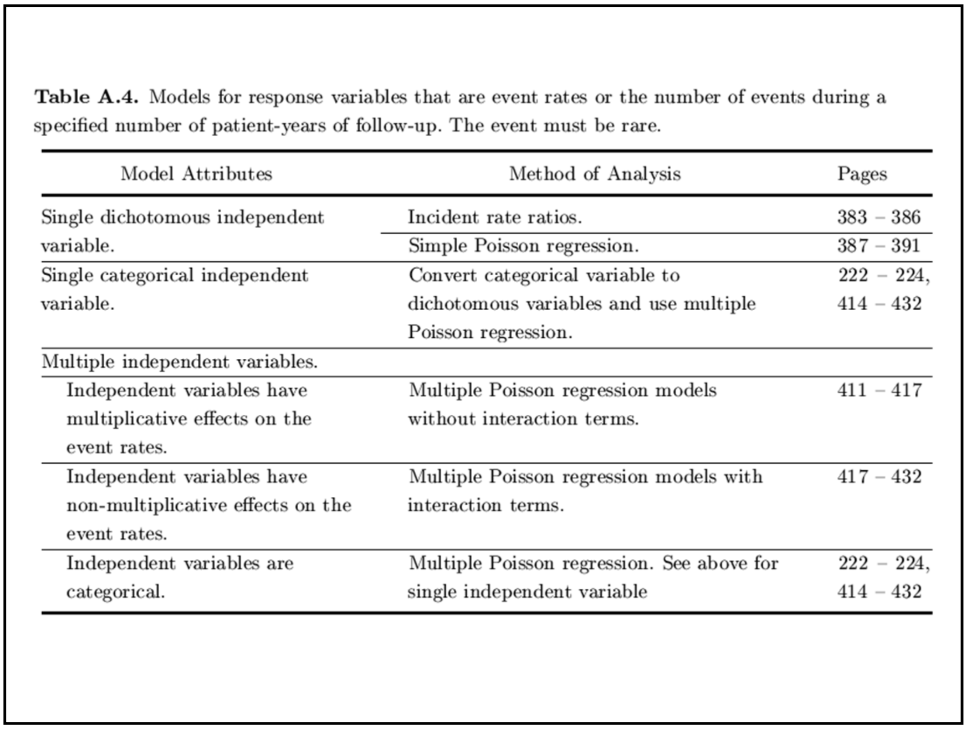


Table A.4. Models for response variables that are event rates or the number of events during a specified number of patient-years of follow-up. The event must be rare.

Model Attributes	Method of Analysis	Pages
Single dichotomous independent variable.	Incident rate ratios.	383 – 386
	Simple Poisson regression.	387 – 391
Single categorical independent variable.	Convert categorical variable to dichotomous variables and use multiple Poisson regression.	222 – 224, 414 – 432
Multiple independent variables.		
Independent variables have multiplicative effects on the event rates.	Multiple Poisson regression models without interaction terms.	411 – 417
Independent variables have non-multiplicative effects on the event rates.	Multiple Poisson regression models with interaction terms.	417 – 432
Independent variables are categorical.	Multiple Poisson regression. See above for single independent variable	222 – 224, 414 – 432

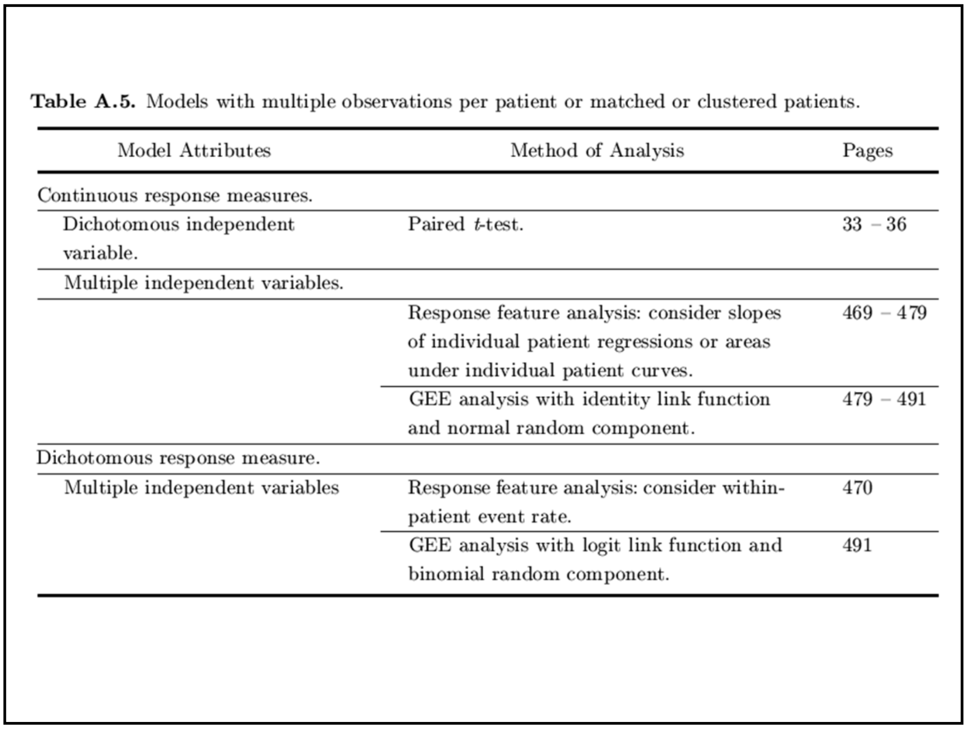
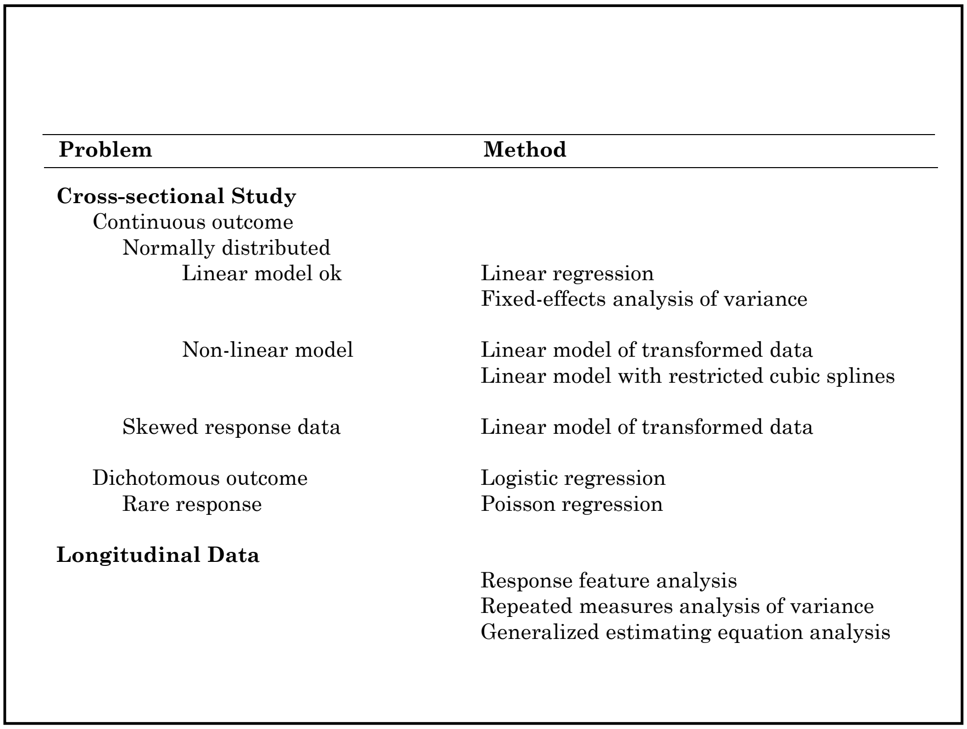
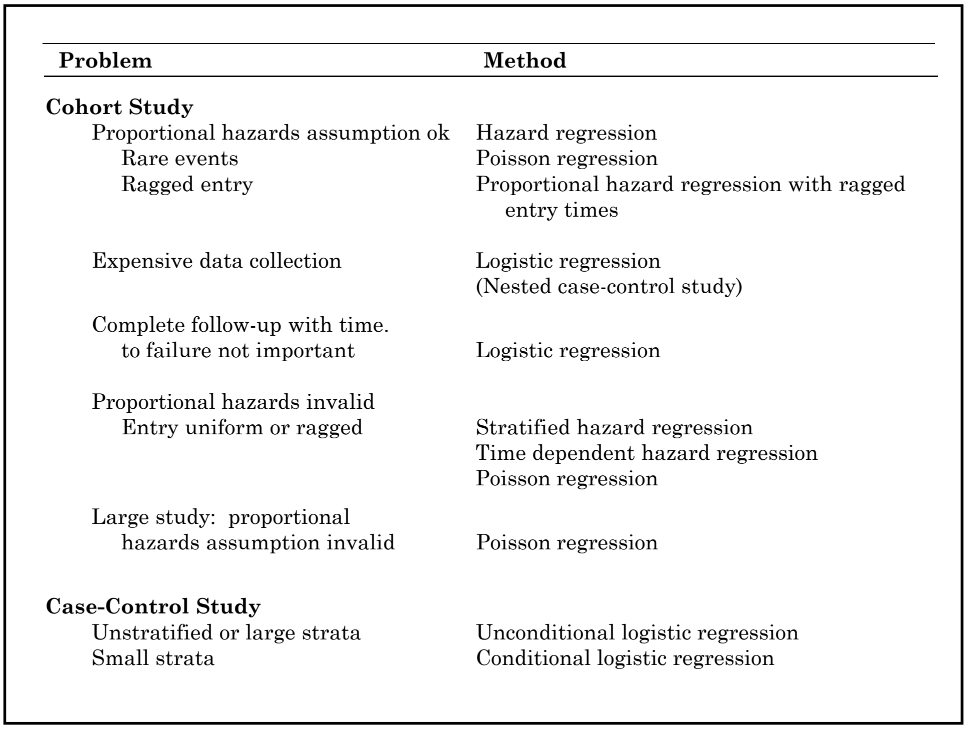


Table A.5. Models with multiple observations per patient or matched or clustered patients.

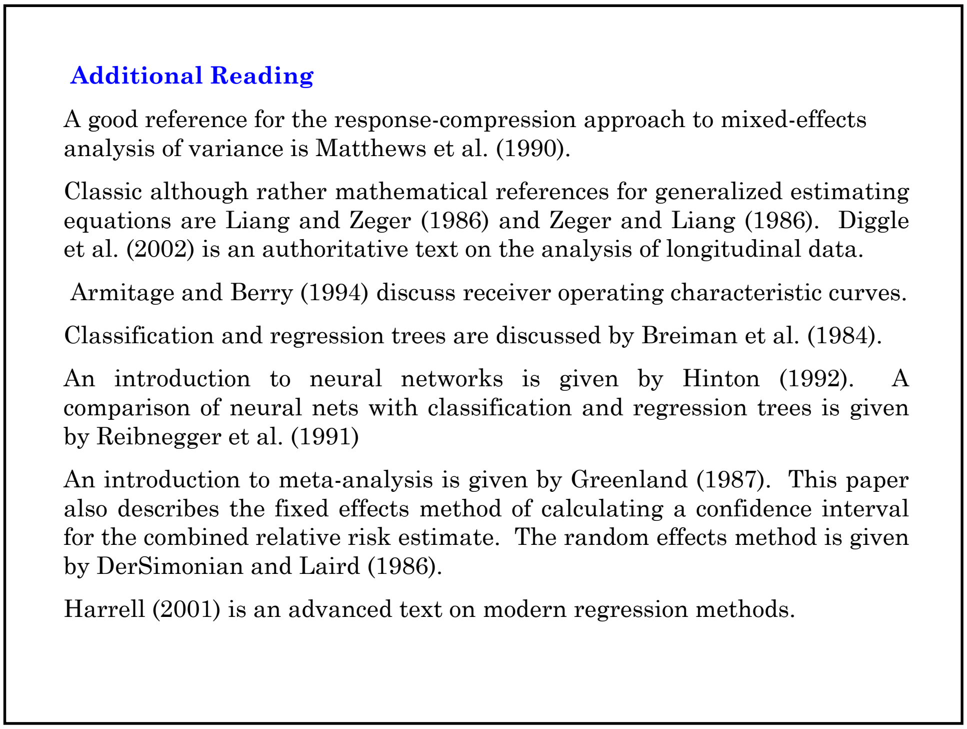
Model Attributes	Method of Analysis	Pages
Continuous response measures.		
Dichotomous independent variable.	Paired *t*-test.	33 – 36
Multiple independent variables.		
	Response feature analysis: consider slopes of individual patient regressions or areas under individual patient curves.	469 – 479
	GEE analysis with identity link function and normal random component.	479 – 491
Dichotomous response measure.		
Multiple independent variables	Response feature analysis: consider within-patient event rate.	470
	GEE analysis with logit link function and binomial random component.	491



Problem	Method
Cross-sectional Study	
Continuous outcome	
Normally distributed	
Linear model ok	Linear regression Fixed-effects analysis of variance
Non-linear model	Linear model of transformed data Linear model with restricted cubic splines
Skewed response data	Linear model of transformed data
Dichotomous outcome	Logistic regression
Rare response	Poisson regression
Longitudinal Data	
	Response feature analysis Repeated measures analysis of variance Generalized estimating equation analysis



Problem	Method
Cohort Study	
Proportional hazards assumption ok	Hazard regression
Rare events	Poisson regression
Ragged entry	Proportional hazard regression with ragged entry times
Expensive data collection	Logistic regression (Nested case-control study)
Complete follow-up with time. to failure not important	Logistic regression
Proportional hazards invalid	
Entry uniform or ragged	Stratified hazard regression Time dependent hazard regression Poisson regression
Large study: proportional hazards assumption invalid	Poisson regression
Case-Control Study	
Unstratified or large strata	Unconditional logistic regression
Small strata	Conditional logistic regression



Additional Reading

A good reference for the response-compression approach to mixed-effects analysis of variance is Matthews et al. (1990).

Classic although rather mathematical references for generalized estimating equations are Liang and Zeger (1986) and Zeger and Liang (1986). Diggle et al. (2002) is an authoritative text on the analysis of longitudinal data.

Armitage and Berry (1994) discuss receiver operating characteristic curves.

Classification and regression trees are discussed by Breiman et al. (1984).

An introduction to neural networks is given by Hinton (1992). A comparison of neural nets with classification and regression trees is given by Reibnegger et al. (1991)

An introduction to meta-analysis is given by Greenland (1987). This paper also describes the fixed effects method of calculating a confidence interval for the combined relative risk estimate. The random effects method is given by DerSimonian and Laird (1986).

Harrell (2001) is an advanced text on modern regression methods.

References

Armitage P and Berry G: *Statistical Methods in Medical Research*, Third ed. Cambridge, MA: Blackwell Science, Inc., 1994.

Bernard GR, Wheeler AP, Russell JA, Schein R, Summer WR, Steinberg KP, Fulkerson WJ, Wright PE, Christman BW, Dupont WD, Higgins SB, Swindell BB: "The Effects of Ibuprofen on the Physiology and Survival of Patients with Sepsis". *New England Journal of Medicine,* 1997; 336: 912-918

Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Belmont CA: Wadsworth, 1984.

Breslow NE and Day NE: *Statistical Methods in Cancer Research: Vol. I The Analysis of Case-Control Studies*. Lyon: IARC Scientific Publications, 1980.

Breslow NE and Day NE: *Statistical Methods in Cancer Research: Vol. II. The Design and Analysis of Cohort Studies*. Lyon: IARC Scientific Publications, 1987.

Cleveland WS. *The Elements of Graphing Data*: Monterey, CA: Wadsworth Advanced Books and Software, Bell Telephone Laboratories, Inc., 1985.

DerSimonian R. Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; 7:177-188.

Diggle PJ, Heagerty P, Liang K-Y, Zeger SL: *Analysis of Longitudinal Data 2nd Ed.* Oxford: Oxford University Press, 2002

Fleiss JL: *Statistical Methods for Rates and Proportions, Second ed.:* New York: John Wiley & Sons, Inc., 1981.

Greene J and Touchstone J. "Urinary Tract Estriol: An Index of Placental Function. *American Journal of Obstetrics and Gynecology*, 1963; 85: 1-9.

Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiologic Reviews* 1987; 9:1-30.

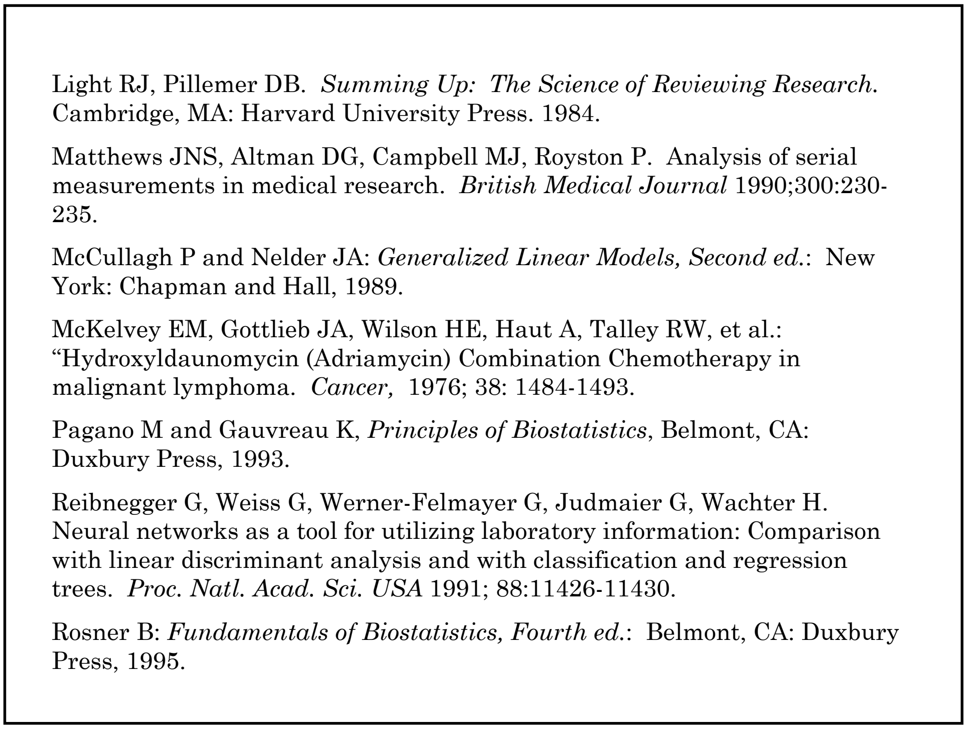
Harrell, FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* New York: Springer, 2001

Hinton GE. How neural networks learn from experience. *Scientific American* September 1992; p.145-151.

Kalbfleisch JD and Prentice RL: *The Statistical Anallysis of Failure Time Data*, New York: John Wiley and Sons, 1980.

Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; 73:13-22.

Liang K-Y, Zeger SL. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986; 42:121-130.



Light RJ, Pillemer DB. *Summing Up: The Science of Reviewing Research.* Cambridge, MA: Harvard University Press. 1984.

Matthews JNS, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. *British Medical Journal* 1990;300:230-235.

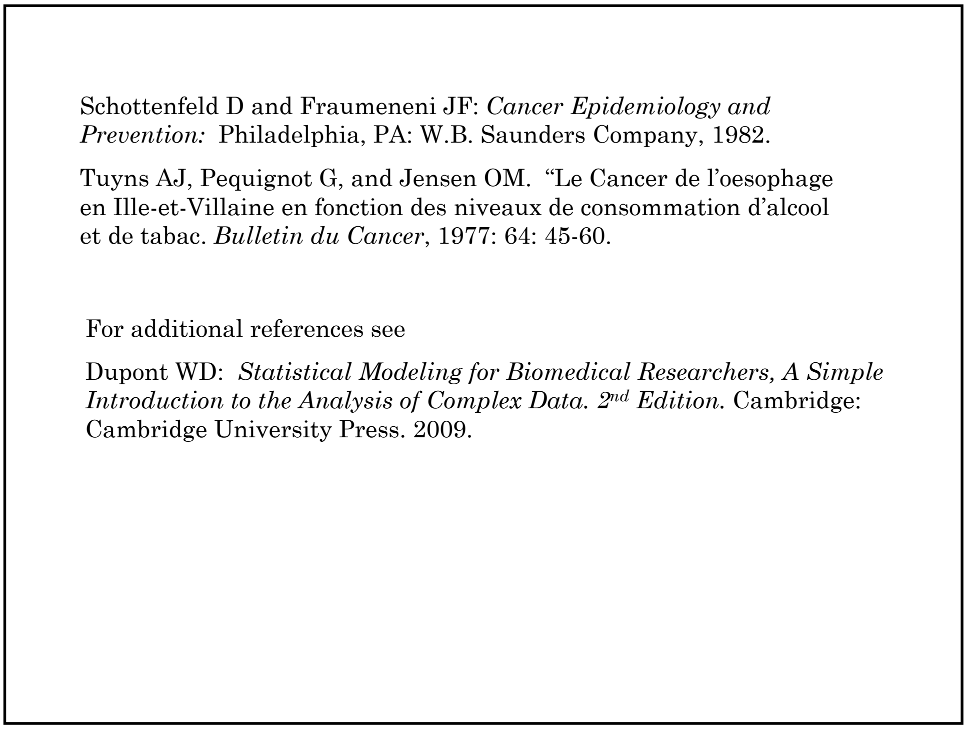
McCullagh P and Nelder JA: *Generalized Linear Models, Second ed.*: New York: Chapman and Hall, 1989.

McKelvey EM, Gottlieb JA, Wilson HE, Haut A, Talley RW, et al.: "Hydroxyldaunomycin (Adriamycin) Combination Chemotherapy in malignant lymphoma. *Cancer,* 1976; 38: 1484-1493.

Pagano M and Gauvreau K, *Principles of Biostatistics*, Belmont, CA: Duxbury Press, 1993.

Reibnegger G, Weiss G, Werner-Felmayer G, Judmaier G, Wachter H. Neural networks as a tool for utilizing laboratory information: Comparison with linear discriminant analysis and with classification and regression trees. *Proc. Natl. Acad. Sci. USA* 1991; 88:11426-11430.

Rosner B: *Fundamentals of Biostatistics, Fourth ed.*: Belmont, CA: Duxbury Press, 1995.



Schottenfeld D and Fraumeneni JF: *Cancer Epidemiology and Prevention:* Philadelphia, PA: W.B. Saunders Company, 1982.

Tuyns AJ, Pequignot G, and Jensen OM. "Le Cancer de l'oesophage en Ille-et-Villaine en fonction des niveaux de consommation d'alcool et de tabac. *Bulletin du Cancer*, 1977: 64: 45-60.

For additional references see

Dupont WD: *Statistical Modeling for Biomedical Researchers, A Simple Introduction to the Analysis of Complex Data. 2nd Edition.* Cambridge: Cambridge University Press. 2009.