Research Methods

Standardization as a Tool for Causal Inference in Medical Research

Safoora Gharibzadeh PhD[1], Kazem Mohammad PhD[1], Abbas Rahimiforoushani PhD[1], Atieh Amouzegar MD[2], Mohammad Ali Mansournia MD MPH PhD•[1]

Abstract

Traditional standardization methods have been used in medical research for a long time to standardize the effect of interest for one confounder such as age. Model-based standardization extension of these methods is used when we have more than one variable produces an effect which is the population average and has marginal causal interpretation.

In this paper, we discuss the most traditional model-based standardization methods that are used to estimate the marginal causal effect of exposure. We applied these methods to data from Tehran Thyroid Study and estimated the standardized effect of exposure on outcome.

Based on the simulation studies, covariate standardization is preferred except when 1) we have enough information about the mechanism of exposure or 2) the outcome is rare and exposure is frequent, so propensity score standardization is suggested.

Keywords: Marginal effect, model-based standardization, parametric g-formula, propensity scores

Cite this article as: Gharibzadeh S, Mohammad K, Rahimiforoushani A, Amouzegar A, Mansournia MA. Standardization as a tool for causal inference in medical research. *Arch Iran Med*. 2016**; 19(9):** 666 – 670.

Introduction

Observational studies are common in epidemiology. These studies generally involve comparing the risk for groups of subjects experiencing different levels of exposure. There is a possibility that an important factor that affects the distribution of outcome may differ between the comparison groups. It is said that the effect of exposure is confounded because this effect is partially due to the difference between groups.[1] More formally, the effect of exposure on outcome is distorted due to a third factor that is related to both exposure and outcome.

Common methods to control confounding are regression (logistic, Poisson regression, and Cox regression, i.e. generalized linear models), stratification, and standardization.

Standardization methods have been used in medical research for a long time.[2–4] These methods are used to control confounding, and one can estimate causal parameters of interest using standardization.[5–7]

Why do we choose standardization to control confounding?

Stratification analysis compares the risks between exposure groups within strata. A consequence of this method is that each individual stratum may contain too little data to be informative.[1] With the increase in the number of confounders, this method is prone to sparse-data problem (it occurs when there are few or no study participants at some combinations of the outcome, exposure, and covariates)[8–10] and unstable estimates. Because of this problem, it is rarely used in practice.

Authors' affiliations: [1]Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, I. R. of Iran, [2]Endocrine Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, I. R. of Iran.

•Corresponding author and reprints: Mohammad Ali Mansournia MD MPH PhD, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, I. R. of Iran. Tel: +98-21-42933046, E-mail: Mansournia_ma@yahoo.com.

Accepted for publication: 3 July 2016

Regression analysis overcomes the stated problem in stratification analysis. Regression analysis is used to model the relationship between confounding variables, exposure and outcome. Modeling binary outcome is usually achieved through logistic regression, which reports odds ratio as an effect measure. The odds ratio is known to be non-collapsible, i.e. in the absence of confounding, the stratum-specific odds ratio (OR) does not equal the crude OR.[11–14] Also, the odds ratio is biologically interpretable only when it is an estimate of the risk ratio or rate ratio. Moreover, logistic regression is not suggested when the outcome is rare and there are many confounding variables.[15]

Limitations of conventional regression analysis

I. These methods produce conditional effect measure, not population average. In most studies, researchers have a marginal or population average treatment effect in mind.[16] Conditional models report the individual level effects while marginal models report population level effects. The intervention level for many exposures such as those including environmental exposures (e.g. air pollution) is population.[17] In this situation, we should use marginal causal models and report population-level estimates.

II. When the causal effect of an exposure on an event varies across the levels of a third variable, we have interaction or effect modification.[18] The traditional statistical methods for controlling confounding cannot handle the situation where the aim is to estimate the effect of exposure on the outcome of interest in the presence of variables which are simultaneously confounders and effect modifiers, because the main assumption in reporting adjusted effect estimates using traditional methods is no interaction between the exposure and confounders.[19] Marginal causal estimators have causal interpretations (indicating that outcome is the result of the occurrence/ presence of the exposure) for the total population even in the presence of interaction.[20]

III. In a longitudinal study, with repeated measurements on the exposure and covariates, the covariates could be confounders for next exposure and simultaneously intermediates for the exposure

of the previous time. In this situation, standard approaches for adjustment of confounding are biased.[21–23]

Standardization using stratification

Standardization involves the calculation of numbers of expected events (e.g. disease/death) which are compared to the number of observed events. With total population as reference, the exposure-specific standardized risks are computed as weighted-average of risk across strata of C with weights equal to the proportion of individuals in each stratum of C.[19]

Suppose that outcome and exposure are binary, and individuals are either exposed or unexposed. In addition to depending on exposure, risks depend on a binary confounder:

Calculating standardized Risk Ratio

Standardized risk in the exposed equals to the probability of disease in different levels of third variable (C) weighted by the distribution of C (the weights equal to $\frac{900}{2000}$ and $\frac{1100}{2000}$ for C = 1 and C = 0, respectively), the calculation for standardized risk in the unexposed was done similarly.

$$\text{Standardized Risk in the exposed} = \left\{\frac{194}{800} * \frac{900}{2000} + \frac{6}{200} * \frac{1100}{2000}\right\} = 0.1256$$

$$\text{Standardized Risk in the unexposed} = \left\{\frac{24}{100} * \frac{900}{2000} + \frac{26}{900} * \frac{1100}{2000}\right\} = 0.1238$$

$$\text{Standardized RR} = \frac{0.1256}{0.1238} = 1.01$$

Model-Based Standardization (MBS) or Covariate Standardization

Generally, in standardization, we compute the risk in exposed and unexposed groups in different levels of confounder and then weigh these risks using the proportion of individuals in each level of C.

Because traditional standardization requires stratification by confounders, the sparse-data problem will occur when stratified by many confounders, and then one might achieve an unstable estimate.[24] To overcome this instability, model-based standardization methods have been proposed.[25,26] If one uses a correct statistical model, MBS can estimate a standardized, or unconfounded, population-averaged effect.[11,27]

When the number of confounder levels increase, or there is more than one confounder some of which are continuous, we cannot tabulate the data and we have to use regression modeling. Model-based standardization usually starts with regular regression modeling. Using regression modeling, we estimate the risk if all participants were exposed or all of them were unexposed and then average these risks over the distribution of confounders.

Parametric g-formula or covariate standardization is the generalization of standardization for time-varying exposures and confounders.[28] This method relies on the same assumptions (no unmeasured confounding, no measurement error and no model misspecification) as alternative methods in causal inference.[29]

Parametric g-formula has three steps. The first step is expansion, in which the data is copied three times. Each individual is simultaneously considered as treated, untreated and in its own treatment status. Then, one should run a suitable regression model (Linear/ Logistic) for the treatment and confounders on the original data set. The third step is prediction, using the parameter estimates from the regression model, one can predict the outcome values for each treated and untreated with the covariates equal to the covariates in the original data, and average over the L to compute mean outcome/ risk in the exposed and unexposed groups. i.e. $\sum_l \hat{E}(D \mid E=1, L=l)$ for exposed and $\sum_l \hat{E}(D \mid E=0, L=l)$ for unexposed.

The overall representation of MBS is as follows:

$$\frac{\sum_l \hat{E}(D \mid E=e, L=l)}{n}$$

Where L is a set of confounders that we should adjust for them, and $\mathrm{E}(D \mid E=e, L=l)$ is the general form of the regression. When the outcome is binary (e.g. logistic regression), the expectation equals probability.

Standardization to Propensity Scores (PS)

Rosenbaum and Rubin developed propensity score methods to make causal inferences in observational data.[7] Propensity score is the conditional probability of receiving treatment (or being exposed) given the observed covariates.[30] Methods based on the propensity score (PS) have become a common approach in causal inference and medical research.[7,31–51]

Assuming that there is no unmeasured confounding, by adjustment for the propensity scores, one can achieve an unbiased estimate of the treatment/exposure effect. Adjustment for the propensity score is typically done through matching, stratification, inverse probability of treatment weighting, and covariate adjustment.[32,47,52]

As an alternative to covariate standardization, propensity score standardization was recently proposed by Hernán and Robins. The general strategy is the same as MBS except that in outcome modeling, L is replaced by propensity score. PS-standardization is a semi-parametric standardization that uses the total group (exposed + unexposed) as the standard.

After estimating propensity score using logistic regression, we compute the population-average risk difference by standardizing the conditional expectation of the outcome to the empirical distribution of propensity score:

$$\frac{\sum_s \hat{E}(D \mid E=e, PS=s)}{n}$$

Where $E(D/E=e_1 PS=s)$ is estimated using logistic regression with exposure and propensity score as covariates. In both methods, confidence interval for effect measure of interest is computed using non-parametric bootstrap with n = 200 bootstrapped samples.

Case Study

To illustrate application of these standardization methods with real data, we studied the effect of categorized waist circumference (WC) (cm) on the incident of diabetes after 10 years of follow-up (median follow up time was 8.7 years) in the presence of these confounders: hypertension, hyperlipidemia, body mass index (BMI) (kg/m2), age (years), sex, and education. We excluded pre-diabetes and diagnosed diabetes at baseline. Pre-diabetic patients were defined as those who had FBG between 100 mg/dL and 125.9 mg/dL.

This study was conducted within the framework of Tehran

Table 1. A population with Exposure E, Disease D, and confounding variable C

Exposure	C = 1			C = 0			Pooled/Crude		
	D = 1	D = 0	Total	D = 1	D = 0	Total	D = 1	D = 0	Total
E = 1	194	606	800	6	194	200	200	800	1000
E = 0	24	76	100	26	874	900	50	950	1000
Total	**RR = 1.01**			**RR = 1.04**			**RR = 4.0**		

Table 2. Cross-tabulation of Exposure and Disease

	Diabetic	Non-Diabetic
Exposed (WC ≥ 95)	50	623
Unexposed (WC < 95)	60	2246

Table 3. Standardized Risk Difference

Method	RD	95% CI
Parametric g-formula	0.0020	0.0004–0.0036
PS-standardization	0.0133	-0.0108–0.0374

Thyroid Study (TTS). It is a prospective population-based cohort study, performed on the residents of district 13 of Tehran with the aim of evaluating the prevalence and natural course of thyroid diseases and their long term consequences in terms of ischemic heart disease and cardiovascular and all-cause mortality in the urban, iodine sufficient population of Tehran, the capital of Iran. Details of the study methods have been previously described.[53, 54]

Variables

Exposure was defined as categorized WC with 95 cm as the cut-off point. The outcome (diabetes) was defined as fasting blood glucose (FBG) ≥ 126 mg/dL or receiving treatment for diabetes.

Diagnosed hypertension was defined as systolic blood pressure (SBP) ≥ 130 mmHg or diastolic blood pressure (DBP) ≥ 85 mmHg, or previously receiving specific treatment. Hyperlipidemia was considered as total cholesterol (TC) greater than 200 mg/dL. Education has three categories: primary, secondary, and higher.

Results

$$Crude.RD = \frac{50}{673} - \frac{60}{2306} = 0.0482$$

Average risk in the exposed equals 0.03811 and 0.04543 using parametric g-formula and propensity score standardization, respectively. Average risk in the unexposed equals 0.03604 and 0.03209 using parametric g-formula and propensity score standardization, respectively.

When we have a rare outcome (prevalence less than 5%) and exposure is frequent, the estimated risk difference from PS-standardization is more reliable.[15] So, in this example, we report standardized RD equal to 0.0133. The authors tested this situation in another simulation study that compared these two standardized estimates in different scenarios.

Discussion

In this paper, we have explained model-based standardization methods. If we want to compare the health status of two populations and we have one or two categorical confounding variables with few levels, we suggest traditional standardization without modeling. When there are many variables and some of them are continuous, model-based standardization should be used.

Overall, based on the results of different studies and the authors' simulation study, covariate standardization works better in different situations with two exceptions: 1) when we have enough information about the mechanism of exposure, such as in pharmaco-epidemiology settings, where the exposure is known by drug indication, PS-standardization is preferred to covariate standardization; 2) when the outcome is rare and exposure is frequent, exposure modeling (PS-standardization) is suggested, because in this situation, there may be too little information to estimate the relationship between outcome and pre-treatment variables, but plenty of data to estimate the relationship between treatment assignment and these variables.[15]

In a point-treatment study, control for confounding is traditionally accomplished by modeling the probability of outcome as a function of treatment and pre-treatment covariates. With a time-varying exposure, traditional methods such as GEE and time-dependent Cox regression may be biased if time-varying covariates are simultaneously confounders and intermediates.[21,22] For example, if we consider the time-varying effect of physical activity on knee pain in patients with osteoarthritis, measurement of BMI before physical activity is a confounder and the BMI after physical activity is a mediator for this effect.[55]

Binary outcome is common in randomized control trials and cohort studies. If the risk ratio is the effect measure of interest, using logistic regression, one can report the odds ratio as a measure of association and use it as approximation of risk ratio when the outcome is rare.[56] Risk ratio is known to be collapsible and easy to interpret. When the outcome is common (> 10%), the adjusted odds ratio derived from the logistic regression can no longer approximate the risk ratio of interest.[57] There are several alternatives that are suggested for modeling the adjusted risk ratio such as Log-binomial regression in this situation.[58] But this model has convergence problem; using standardization methods that are described in this article, one can estimate the risk in the exposed and in the unexposed, and then compute the risk ratio. However, we should notice that this risk ratio is a marginal effect, but using Log-binomial regression, we achieve the conditional risk ratio.

Recommendations for researchers

When our aim is to estimate population average effect of exposure/treatment, model-based standardization is one of the best approaches, which can be done in two ways: covariate standardization and propensity score standardization. In the first step, the researchers should check the prevalence of exposure and outcome in the population and then if the outcome is rare and exposure is frequent, they should report the result of propensity score standardization, otherwise covariate standardization is preferred. These methods can be used in prospective studies, such as cohort and RCT.

References

1. Clayton D, Hills M, Pickles A. Statistical models in epidemiology. *IEA*. 1993; 135 – 145.
2. Breslow NE, Day NE. *Statistical methods in cancer research*. International Agency for Research on Cancer Lyon, 1987.
3. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Lippincott Williams & Wilkins, 2008.
4. Keiding N, Clayton D. Standardization and control for confounding in observational studies: a historical perspective. *Statistical Science*. 2014; 29: 529 – 558.
5. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology*. 1986; 15: 413 – 419.
6. Miettinen O. Components of the crude risk ratio. *American Journal of Epidemiology*. 1972; 96: 168 – 172.
7. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70: 41 – 55.
8. Josefsson M, de Luna X, Daniels MJ and Nyberg L. Causal inference with longitudinal outcomes and non-ignorable drop-out: Estimating the effect of living alone on cognitive decline. *Journal of the Royal Statistical Society Series C, Applied Statistics*. 2016; 65: 131 – 144.
9. Greenland S, Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Statistics in Medicine*. 2015; 34: 3133 – 3143.
10. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *BMJ*. 2016; 352: i1981.
11. Cai Z and Brumback BA. Model-based standardization to adjust for unmeasured cluster-level confounders with complex survey data. *Statistics in Medicine*. 2015; 34: 2368 – 2380.
12. Cole SR, Lau B, Eron JJ, Brookhart MA, Kitahata MM, Martin JN, et al. Estimation of the standardized risk difference and ratio in a competing risks framework: application to injection drug use and progression to AIDS after initiation of antiretroviral therapy. *American Journal of Epidemiology*. 2014: kwu122.
13. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Statistical Science*. 1999: 29 – 46.
14. Mansournia MA, Greenland S. The relation of collapsibility and confounding to faithfulness and stability. *Epidemiology*. 2015; 26: 466 – 472.
15. Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores. *Annals of Internal Medicine*. 2002; 137: 693 – 695.
16. Martens EP, Pestman WR, Klungel OH. Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: A Monte Carlo study (pn/a) by Peter C. Austin, Paul Grootendorst, Sharon-Lise T. Normand, Geoffrey M. Anderson, Statistics in Medicine, Published Online: 16 June 2006. *Statistics in Medicine*. 2007; 26: 3208 – 3210.
17. Hubbard AE, Van Der Laan MJ. Population intervention models in causal inference. *Biometrika*. 2008; 95: 35 – 47.
18. Kamangar F. Effect modification in epidemiology and medicine. *Arch Iran Med*. 2012; 15(9):575 – 582.
19. Mohammad K, Hashemi–Nazari SS, Mansournia N, Mansournia MA. Marginal versus conditional causal effects. *Journal of Biostatistics and Epidemiology*. 2015; 1: 121 – 128.
20. Hernán MA, Robins JM. *Causal inference*. Boca Raton: Chapman & Hall/CRC, forthcoming, 2016.
21. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000: 550 – 560.
22. Bodnar LM, Davidian M, Siega–Riz AM, Tsiatis AA. Marginal structural models for analyzing causal effects of time–dependent treatments: an application in perinatal epidemiology. *American Journal of Epidemiology*. 2004; 159: 926 – 934.
23. Hernán MÁ, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV–positive men. *Epidemiology*. 2000; 11: 561 – 570.
24. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology*. 2003: 680 – 686.
25. Flanders WD, Rhodes PH. Large sample confidence intervals for regression standardized risks, risk ratios, and risk differences. *Journal of Chronic Diseases*. 1987; 40: 697 – 704.
26. Greenland S. Estimating standardized parameters from generalized linear models. *Statistics in Medicine*. 1991; 10: 1069 – 1074.
27. Zheng HW, Brumback BA, Lu X, Bouldin ED, Cannell MB, Andresen EM. Doubly robust testing and estimation of model-adjusted effect-measure modification with complex survey data. *Statistics in Medicine*. 2013; 32: 673 – 84.
28. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*. 1986; 7: 1393 – 1512.
29. Taubman SL, Robins JM, Mittleman MA, Hernán MA. Intervening on risk factors for coronary heart disease: an application of the parametric g–formula. *International Journal of Epidemiology*. 2009; 38: 1599 – 1611.
30. Hernan MA, Robins JM. *Causal Inference*. Taylor & Francis, 2013.
31. Anstrom KJ, Tsiatis AA. Utilizing Propensity Scores to Estimate Causal Treatment Effects with Censored Time-Lagged Data. *Biometrics*. 2001; 57: 1207 – 1218.
32. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*. 2011; 46: 399 – 424.
33. Becker SO, Ichino A. Estimation of average treatment effects based on propensity scores. *The Stata Journal*. 2002; 2: 358 – 377.
34. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*. 1984; 79: 516 – 524.
35. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Annals of internal Medicine*. 1997; 127: 757 – 763.
36. Rubin DB. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*. 2001; 2: 169 – 188.
37. Rubin DB. On principles for modeling propensity scores in medical research. *Pharmacoepidemiology and Drug Safety*. 2004; 13: 855 – 7.
38. Rubin DB, Thomas N. Matching using estimated propensity scores: Relating theory to practice. *Biometrics*. 1996: 249 – 264.
39. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *Journal of Clinical Epidemiology*. 2005; 58: 550 – 559.
40. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2011; 174: 369 – 386.
41. Bidwell LC, Palmer RH, Brick L, Madden PA, Heath AC, Knopik VS. A propensity scoring approach to characterizing the effects of maternal smoking during pregnancy on offspring's initial responses to cigarettes and alcohol. *Behavior Genetics*. 2016: 1 – 15.
42. d'Agostino RB. Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non–randomized control group. *Stat Med*. 1998; 17: 2265 – 2281.
43. D'Agostino RB. Propensity scores in cardiovascular research. *Circulation*. 2007; 115: 2340 – 2343.
44. Freedman DA, Berk RA. Weighting regressions by propensity scores. *Evaluation Review*. 2008; 32: 392 – 409.
45. Hemmila MR, Birkmeyer NJ, Arbabi S, Osborne NH, Wahl WL, Dimick JB. Introduction to propensity scores: a case study on the comparative effectiveness of laparoscopic vs open appendectomy. *Archives of Surgery*. 2010; 145: 939 – 945.
46. Jonklaas J, Sarlis NJ, Litofsky D, Ain KB, Bigos ST, Brierley JD, et al. Outcomes of patients with differentiated thyroid carcinoma following initial therapy. *Thyroid*. 2006; 16: 1229 – 1242.
47. Lee BK, Lessler J, Stuart EA. Weight trimming and propensity score weighting. *PloS One*. 2011; 6: e18174.
48. Lee J, Park JH, Lee CR, Chung WY, Park CS. Long–term outcomes

of total thyroidectomy versus thyroid lobectomy for papillary thyroid microcarcinoma: comparative analysis after propensity score matching. *Thyroid*. 2013; 23: 1408 – 1415.
49. Månsson R, Joffe MM, Sun W, Hennessy S. On the estimation and use of propensity scores in case–control and case–cohort studies. *American Journal of Epidemiology*. 2007; 166: 332 – 339.
50. Perkins SM, Tu W, Underhill MG, Zhou XH, Murray MD. The use of propensity scores in pharmacoepidemiologic research. *Pharmacoepidemiology and Drug Safety*. 2000; 9: 93 – 101.
51. Steiner PM, Cook D. Matching and propensity scores. *The Oxford Handbook of Quantitative Methods in Psychology*. 2013; 1: 237.
52. Mansournia MA, Altman DG. Inverse probability weighting. *BMJ*. 2016; 352: i189.
53. Azizi F, Amouzegar A, Delshad H, Tohidi M, Mehran L. Natural course of thyroid disease profile in a population in nutrition transition: Tehran Thyroid Study. *Arch Iran Med*. 2013; 16: 418.
54. Amouzegar A, Beigy M, Gharibzadeh S, Azizi F. Underestimation of thyroid dysfunction risk due to regression dilution bias in a long–term follow–up: Tehran Thyroid Study (TTS). *Horm Metab Res*. 2014; 46: 440 – 444.
55. Mansournia MA, Danaei G, Forouzanfar MH, Mahmoodi M, Jamali M, Mansournia N, et al. Effect of physical activity on functional performance and knee pain in patients with osteoarthritis: Analysis with marginal structural models. *Epidemiology*. 2012; 23: 631 – 640.
56. Janani L, Mansournia MA, Nourijeylani K, Mahmoodi M, Mohammad K. Statistical Issues in Estimation of Adjusted Risk Ratio in Prospective Studies. *Arch Iran Med*. 2015; 18(10): 713 – 719.
57. Zhang J, Kai FY. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *Jama*. 1998; 280: 1690 – 1691.
58. Wacholder S. Binomial regression in GLIM: Estimating risk ratios and risk differences. *American Journal of Epidemiology*. 1986; 123: 174 – 184.