Developing Scientific Research Proposals (Grant Writing)

2003 Epidemiology and Biostatistics Summer Session

Alan Kristal, Dr.P.H.

Session 6

Statistical Analysis

Content of Analysis Section

- **Articulate the hypotheses that correspond to each specific aim**
- **Specify key variables**
- **Describe the statistical model**
- **Describe the coefficient(s) or statistic(s) that test(s) each specific aim**

Content of Analysis Section

Articulate each hypothesis

"*Specific Aim 1* examines the associations of dietary fat with risk of breast cancer recurrence."

Content of Analysis Section

Articulate each hypothesis

"Analyses will examine whether treatment with Cox-2 inhibitors is associated with increased apoptosis (*Specific Aim 1*) and decreased cell proliferation rate (*Specific Aim 2*) in tissues obtained from prostate needle biopsy."

Content of Analysis Section

Specify key variables

"Definition of Cancer Recurrence: For the primary analysis, breast cancer recurrence will be defined as the time from randomization to diagnosis. Participants without a breast cancer recurrence will be censored at the time of their last follow-up phone contact. Participants lost to follow-up will be censored at the time of their last follow-up."

Content of Analysis Section

Specify key variables

"In all analyses, treatment with Cox-2 inhibitors will defined by randomized treatment assignment (100mg/d, 50mg/d, or placebo)."

Content of Analysis Section

Specify key variables

"Definition of Exposure Variables: The three primary exposures are the average intake per day of omega-3 fatty acids, linoleic acid, and total saturated fatty acids. Average intake of these nutrients will be calculated as the arithmetic mean of total daily intake (foods plus supplements) from six, 24-hour dietary recalls."

Content of Analysis Section

Describe the statistical model

"The relative risk of lung cancer associated with beta-carotene supplementation will be estimated using Cox proportional hazards models. The form of these models will be…."

Content of Analysis Section

Describe the statistical model

"To determine how many factors will be examined, we will use the scree plot from a principal components factor analysis, specifying squared multiple correlation in the diagonal of the identity matrix. If there is no clear breakpoint in the scree plot, we will examine a set of possible solutions based on the numbers of factors around which the scree plot crosses eigenvalues of 1.0. To assist interpretation of each set of potential factor solutions, we will use a non-orthogonal rotation."

Content of Analysis Section

Describe the coefficient or statistic that tests the hypothesis

"In this model, the exponentiated regression coefficient (e^{β}) for the variable indicating history of amphetamine use (coded 0,1) gives the relative risk of pharyngeal cancer associated with amphetamine use. The standard error of the regression coefficient will be used to calculate the 95% confidence interval around this risk."

Content of Analysis Section

Describe the coefficient or statistic that tests the hypothesis

"The treatment effect will be calculated as the change in weight among the intervention group minus change in weight among controls. This value will be estimated from a regression model in which the independent variable is a dummy variable (coded 0,1) indicating treatment arm. The coefficient for the treatment indicator variable and its standard error will be used to test the hypothesis that treatment decreases weight gain over the 36 month observation period."

Overview: Statistical Models

Standard Techniques for Analysis of Epidemiologic Studies

Dependent Variables	Statistical Technique
Continuous	Differences in means (ANOVA)
	Correlation/Regression
Dichotomous	Difference in percentages
	Odds Ratios/Relative Risks
	Logistic regression

Overview: Statistical Models

Standard Techniques for Analysis of Epidemiologic Studies

Dependent Variables	Statistical Technique
Ordered Categories	Trends in percentages or odds ratios
	Polychotomous logistic regression (assume proportional increase)
Multiple Categories	Differences among percentages
	Odds ratios, common contrast group
	Polychotomous logistic regression (no proportional relationships)

Overview: Statistical Models

Standard Techniques for Analysis of Epidemiologic Studies

Dependent Variables	Statistical Technique
Time to failure	Life tables
	Proportional hazards

Parameterizaton of Statistical Models

"We will create indicator variables (x_1, x_2, x_3) for three tertiles of exposure to ultraviolet radiation. Tertiles will be defined by exposure in controls."

"We will log-transform total fat intake (in grams), to normalize its distribution."

Parameterizaton of Statistical Models

"Change in weight will be calculated as change in kilograms from baseline to the 12-month follow-up."

"Percentage energy from fat will calculated as [fat(g)*9/energy (kcal)*100. Energy will include total macronutrients including alcohol."

Parameterizaton of Statistical Models

"Case-control status will be coded as follows: 0=no cancer; 1=stage A and B; and 2=stage C and D."

Special Analysis Issues

Missing Data

- Eliminate from analyses
- Sensitivity analyses
- Imputation

Undetectable laboratory values

- Assign midrange between zero and lowest detectable value

Special Analysis Issues

Outliers

- Eliminate from analyses
- Truncate to lowest or highest acceptable value

Drop Outs/Lost to Follow Up/Death

- Censor
- Eliminate from analyses
- Sensitivity analyses

Analysis Variations

Different forms of parametizing exposures

- Continuous vs. ordered categories
- Scales vs. individual items
- Different quantification of exposure (as raw value, percentage of total, change from baseline)
- Alternative transformations
- Interactions (categorical vs. continuous)
- Time of endpoint (early vs. late effects)

Analysis Variations

Different covariate sets

- Intervention effects with and without covariates
- Covariates that may be mediators of effects
- Highly intercorrelated variables

Analysis Variations

Address “hot” topics and controversies

Nutritional epidemiology:

- Models for energy adjustment
- Exclusion of energy outliers
- Adjustment for bias and reliability

Analysis Variations

Address “hot” topics and controversies

Cancer epidemiology:

- Stage/grade of disease
- Molecular characterization of tumor
- Screening

Analysis Variations

Different definition of study population

- Stratification by demographic characteristics
- Stratification by disease subtypes
- Exclusion of early cases

Analysis Section - Strategies

Begin with overview paragraph

- Write as an analysis “abstract”
- Give overview that orients critical readers yet can be understood by non-statisticians

Analysis Section - Strategies

Justify all decisions!!!!

Explain all critical choices in statistical models, definition of variables, specification of covariates, etc.

Analysis Section - Strategies

Too many analyses in Primary Aims?

- Give details for one or two as examples
- Select examples that illustrate your competence and understanding of complex issues

Analysis Section - Strategies

Secondary Aims

- Articulate the analysis strategy
- Give modest amount of detail
- Develop fully only if you have the space

Analysis Section - Strategies

Exploratory analyses

- Describe additional analyses that are related to but do not test specific aims
- Impress reviewer with scientific depth of proposed study
- Demonstrate your understanding and competence to fully utilize all data collected