Published in final edited form as:
Prev Sci. 2003 September ; 4(3): 155–171.

Advances in Statistical Methods for Substance Abuse Prevention Research

David P. MacKinnon[1,2] and **Chondra M. Lockwood**[1]

[1]Department of Psychology, College of Liberal Arts & Sciences, Arizona State University, Tempe, Arizona.

Abstract

The paper describes advances in statistical methods for prevention research with a particular focus on substance abuse prevention. Standard analysis methods are extended to the typical research designs and characteristics of the data collected in prevention research. Prevention research often includes longitudinal measurement, clustering of data in units such as schools or clinics, missing data, and categorical as well as continuous outcome variables. Statistical methods to handle these features of prevention data are outlined. Developments in mediation, moderation, and implementation analysis allow for the extraction of more detailed information from a prevention study. Advancements in the interpretation of prevention research results include more widespread calculation of effect size and statistical power, the use of confidence intervals as well as hypothesis testing, detailed causal analysis of research findings, and meta-analysis. The increased availability of statistical software has contributed greatly to the use of new methods in prevention research. It is likely that the Internet will continue to stimulate the development and application of new methods.

Keywords

prevention; statistical methods; substance abuse

OVERVIEW

There are many innovative and exciting statistical methods now being developed and applied to substance abuse prevention research data. Standard statistical analyses are often inadequate in substance abuse prevention research because of the special characteristics of the data collected in these studies. Categorical as well as continuous measures, clustering of individuals in schools and clinics, repeated measurements, and missing data are common in substance abuse research data. Statistical methods to address these and other unique aspects of substance abuse research data are the focus of this paper. Citations describing the methods and software for the analyses are discussed in the text and summarized in Table 1 along with sources of information on the Internet.

Standard Model for the Analysis of Program Effects

The most common model for the estimation of program effects on a continuous dependent measure is the conditional linear regression model in Eq. 1.

2Correspondence should be directed to David P. MacKinnon, PhD, Department of Psychology, Arizona State University, Tempe, Arizona 85287-ll04; davidpm@asu.edu.

$$Y_{ij}=\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+e_{ij} \quad (1)$$

where Y_{ij} is the dependent measure for the *i*th individual at measurement *j*, β_1 codes the effect of the baseline measure of the dependent variable (X_1), β_2 codes the effect of a prevention program (X_2), β_3 codes the effect of a covariate such as age (X_3), and e_{ij} is the residual error for the *i*th individual at the *j*th measurement. The estimate of a substance abuse program effect is equal to β_2 and a *t* test of the significance of the program effect is obtained by dividing the estimate of β_2 by its estimated standard error. The model assumes randomization of individuals to conditions, independent observations, reliable and valid measures, and complete data. If there are one or more measurements after baseline, the model can be used to estimate program effects at each wave of measurement adjusted for the baseline measurement. A repeated measures version of this model includes multiple dependent variables on the left-hand side of the equation where the additional variables correspond to additional waves of measurement. Estimates of program effects across the waves of measurement and tests of trends are available as contrasts. These contrasts across the repeated measures are used to test alternative hypotheses about the program effect in later measurements (Abelson & Prentice, 1997). In particular, the β_2 coefficient for the program effect coded in X_2 can be expanded to include both linear and quadratic trends to model the growth or decay in the program effect over time. Similarly, the effects of booster programming can be addressed with contrasts.

The statistical analysis and interpretation of the model in Eq. 1 is taught in most graduate research programs and is described in several books including Keppel (1991),Kirk (1995), and Cohen and Cohen (1983). Major statistical programming packages such as SAS (SAS, 1999) and SPSS (SPSS, 1999) include routines to conduct the analysis. The statistical power calculations to determine the number of participants needed in a research study to detect program effects of a certain size are also described in these books and software is available to compute power for these designs (Cohen, 1988; Gpower, see Table 1).

Often the dependent variable in substance abuse research is categorical such as a dichotomous variable coding whether or not a participant smoked a cigarette in the last month. When the dependent variable is categorical, several assumptions of the ordinary analysis described above are violated (e.g., the errors, e_{ij}, in Eq. 1 are no longer normal) and a more accurate method is needed. Logistic regression is the most commonly used method to analyze categorical dependent variables where the regression coefficients in Eq. 1 are the logarithm of the odds ratio for a one unit change in the independent variable. Again, the program effect is equal to β_2 and statistical significance is tested by dividing β_2 by its standard error. Typically, logistic regression methods are taught only in graduate programs with an emphasis on epidemiology or categorical data analysis, using books such as Hosmer and Lemeshow (1989) or Selvin (1996). Power to detect odds ratios are described in Hsieh (1989). Several computer programs calculate power for logistic regression analysis (e.g., EGRET, 1991).

Statistical Methods Addressing Unique Aspects of Substance Abuse Research Data

Substance abuse data often has characteristics that make the above models difficult to justify. This section describes several unique aspects of substance abuse research data and the methods to accurately analyze the data.

MULTILEVEL MODELS

Participants in substance abuse research are often clustered within larger units such as schools, clinics, communities, or families (Yin & Kaftarian, 1997). Analyzing clustered data only at the individual level is problematic. In the school case for example, it is likely that the students

in the same school are more similar to each other than they are to students in other schools. Because of this dependency among students in each school, the observations are no longer independent, thus violating the independent observations assumption of many statistical methods.

A formal measure of this dependency is the intraclass correlation (Haggard, 1958; Murray *et al*., 1994). A positive intraclass correlation results in inflated rates of Type I error (falsely rejecting a true null hypothesis) for the regression and logistic regression models (Barcikowski, 1981). Even when the intraclass correlation appears small, it can seriously increase Type I errors. In many early substance use studies, individual data were analyzed even though the data were clustered in schools. As a result, these studies may have erroneously led to the conclusion that a program had a significant effect. Fortunately, a statistical method called multilevel analysis or random coefficient modeling can now appropriately analyze clustered data (Bryk & Raudenbush, 1992; Little *et al*., 2000). Recent applications of this type of analysis to substance abuse prevention are the cross-site evaluation of CSAP's high-risk youth programs (Sambrano, 1996) and CSAP's community partnerships (Yin *et al*., 1997).

Several levels of data can be analyzed simultaneously using multilevel analysis. For the case of individuals in schools, the multilevel model includes models for both the individual (Level 1) and school (Level 2) levels. At Level 1, a linear model is specified for individuals within each school. Parameters in this model are assumed to be random and vary as a function of predictors at the school level. The following equations summarize these relations:

$$\text{Individual Level 1: } Y_{ij}=\beta_{0j}+\beta_{1j}X_{ij}+e_{ij} \tag{2}$$

$$\text{School Level 2: } \beta_{0j}=\gamma_{00}+\gamma_{01}W_j+u_{0j} \tag{3}$$

$$\beta_{1j}=\gamma_{10}+u_{1j} \tag{4}$$

In these equations, subscript *i* refers to individuals and subscript *j* refers to schools. The equations illustrate how both individual-level predictors in the *X* matrix (like age, gender, or other covariates) and school-level predictors in the *W* matrix (like assignment to program or control groups or other school characteristics) can be analyzed simultaneously in a multilevel analysis. The estimation of error terms at both levels of the model (e_{ij} at the individual level and u_{0j} and u_{1j} at the school level) allows for a nonzero intraclass correlation to be incorporated in the analysis.

Besides adjusting for a nonzero intraclass correlation, multilevel models can be used to examine effects at the different levels (e.g., the school effect on individuals). Classroom can be included as a third level in the analysis allowing for the examination of school and classroom effects on individuals. To date, these important cross-level effects have not been addressed in much detail in the drug prevention literature (Krull & MacKinnon, 2001; Palmer *et al.*, 1998).The potential for these cross-level relationships makes it important that researchers include measures of school, classroom, and any other potential clustering variable in their research. The hypotheses about different levels also apply to the clustering of individuals in clinics, families, or therapists.

Power calculations that adjust for the effects of clustering are described in Donner (1985), Barcikowski (1981), and Murray (1998) for both continuous and categorical dependent variables. An Internet site (see Table 1, Prevention Research and Methodology Group) that

can be used to calculate power for multilevel data analysis is also available (Brown *et al.*, 1998).

There are a number of software programs that conduct multilevel analysis. Programs specifically designed for the analysis of multilevel models include HLM (Bryk *et al.*, 1988) and MLn (Goldstein *et al.*, 1998). Both of these programs contain subroutines that accommodate categorical outcome variables (Goldstein, 1991; Goldstein & Rasbash, 1996). Routines for the analysis of continuous or categorical data that are clustered are also available for downloading on Donald Hedeker's website (see Table 1; Hedeker & Gibbons, 1996, 1997). Because the multilevel model is a special case of the general mixed model, the MIXED procedure of the SAS software system (SAS Institute, Inc., 1992) can be used to generate the necessary estimates (Brown & Prescott, 1999; Murray, 1998; Singer, 1998; Willett & Singer, 1998). The SAS MIXED software can be combined with the SAS GLMMIX (Littell *et al.*, 1996) subroutine to conduct multilevel analysis of a categorical dependent variable.

The multilevel model can be extended to conduct a form of geographical data analysis where information on the location of data collection is incorporated in the analysis (Anselin & Florax, 1995). Recent geographic analyses of the relationship between the alcohol establishment density and alcohol-related problems (such as rates of assaultive violence) illustrate this method. Some researchers have argued that the geographic proximity of alcohol outlets violates the independent observations assumption (Gruenewald *et al.*, 1996). The SAS MIXED program includes options to incorporate dependency due to spatial proximity.

LONGITUDINAL MODELS

Longitudinal measurement of substance abuse is important for the study of the development of substance abuse. There has been much methodological development in the analysis of longitudinal data and many researchers have shown the applications of this new technology (e.g., Bryk & Raudenbush, 1992; Duncan *et al.*, 1999; Muthén & Curran, 1997; Muthén & Khoo, 1998; Stoolmiller, 1994), including applications in substance abuse prevention (e.g., Duncan & Duncan, 1996). One of the newest applications is the use of latent growth curve modeling (LGM) in the structural equation modeling framework using LISREL (Jöreskog & Sörbom, 2000), AMOS(Arbuckle, 1997), Mplus (Muthén & Muthén, 1998), or EQS (Bentler, 1997).

LGM is a statistical method to measure, explain, and describe individual differences in change over time (Rogosa, 1988). When data are collected on many individuals over several observations, LGM assesses individual growth or development curves. The growth model consists of two levels of assessment, the repeated observations or within-person model (Level 1) and the person-level or between-person model (Level 2). At Level 1 all members of the population are assumed to have trajectories of the same form (e.g., all quadratic and linear terms), but each individual can have different values for the growth parameters that include initial status (intercept) and rate of change (slope; Willett & Sayer, 1994). These growth parameters then become the outcome variables at Level 2, where they are predicted by person-level characteristics (Bryk & Raudenbush, 1992). The specification of the growth curve in terms of levels of analysis illustrates its similarity to multilevel modeling. In fact, multilevel modeling and covariance structure modeling both are used to analyze growth curve models.

Other latent growth curve methods allow for the appropriate analysis of data that includes repeated measures from several cohorts of participants across several waves. This approach includes a level of analysis for cohort as well as the within-person and between-person levels described earlier. These methods examine convergence across cohorts and ages following procedures outlined by Anderson (1995) and Duncan *et al.* (1996).

With repeated categorical dependent variables, logistic regression analysis is much more complicated because of the difficulty of modeling the associations among the repeated categorical observations. Some of the growth curve methods are available for such data (Muthén, 1998). A method for repeated categorical data called generalized estimating equations (GEE; Brown & Liao, 1999; Diggle *et al*., 1994) has recently been applied to substance use research data (Chou *et al*., 1998). The GEE model is a very general model that appropriately models dependencies in repeated categorical data. The analysis is now included in the SAS program GENMOD making the approach considerably easier to conduct (SAS Institute, Inc., 1997).

Collins and colleagues have developed a procedure called latent transition analysis that generates stage-sequential models of individual categorical growth data (Collins & Wugalter, 1992). The method is an extension of Guttman scaling. Guttman scaling is widely used to test whether drug use progresses from drugs such as tobacco and alcohol to harder drugs such as heroin and cocaine (Kandel, 1991). Latent transition analysis tests hypotheses about the ordering of drug use initiation as well as predictors of the ordering (Hyatt & Collins, 2000). This approach seems ideal to test theories that postulate these unique stages in the progression from addiction to nonaddiction (DiClemente *et al*., 1991).

SURVIVAL ANALYSIS

A focus of some substance abuse research studies is the length of time until an event occurs, for example, time until relapse to drug use. Analysis of the time until an event as the dependent variable in Eq. 1 has several limitations including the lack of a clear way to incorporate data from persons who never experienced the event within the observed time of the study. Survival analysis is a group of statistical methods used to investigate time until an event occurs, such as time until death within the time frame observed, that incorporates information on persons who do not experience the event, called censoring (Hosmer & Lemeshow, 1999). These methods have recently been used in substance abuse prevention to examine onset of drug use (Kosterman *et al*., 2000), recidivism of persons convicted of driving under the influence of alcohol (Voas *et al*., 1999), and attrition from a study (Siddiqui *et al*., 1996). Survival analysis has also been used as a method to handle missing data in a drug prevention study (Bacik *et al*., 1998).

A useful measure from survival analysis, the hazard rate, refers to the likelihood of an event among persons who did not have the event up to a certain time. Voas *et al*. (1999) used survival analysis to demonstrate that the time until DUI rearrest was longer for offenders with an alcohol interlock device (a device that will not allow a car to be driven if the driver's breath contains a certain amount of alcohol) than it was for a group of offenders who did not have the device. The hazard rate, the likelihood of being rearrested for DUI six months after the first DUI conviction, given that the offender had not been rearrested before that time, provided an instantaneous measure of the risk of DUI rearrest. Computer software for survival analysis is now available as part of many programs including SAS (1999) and EGRET (1991).

A new development is the generalization of the multilevel models described earlier to conduct survival analysis (see Table 1). Observations from each individual are considered to be clustered within the individual. Therefore, the number of observations can differ across individuals because of different survival times.

MEASUREMENT

All of the statistical methods described previously assume that measures are reliable and valid. If measures are not reliable, then observed relationships among them may be attenuated according to the degree of unreliability (Crocker & Algina, 1986; Lord & Novick, 1968). If

measures are not valid, then conclusions about relationships among variables are misleading. One of the first advanced methodologies applied to substance abuse prevention was covariance structure modeling (Bentler, 1980; Bentler & Newcomb, 1986; Bollen, 1989), at least in part because of its ability to model the reliability of measures. There are two types of variables in covariance structure modeling, measured variables and latent variables. Latent variables are unobserved, theoretical constructs that cannot be directly measured but must be inferred on the basis of observed variables (Bollen, 2002). Measured variables are observed variables, such as questionnaire items. A measurement model is used to develop more reliable measures of latent variables by specifying how measured variables are related to the latent variable. When multiple indicators are used to measure a latent variable, the latent variable is more reliable than each individual item. The structural model specifies the relationships among these latent factors. Covariance structure models have been applied widely in substance abuse research (Muthén, 1998; Newcomb & Bentler, 1988). There have been considerable improvements in the statistical software to estimate these models (Bentler, 1997; Muthén & Muthén, 1998). Programs suitable for the analysis of categorical variables include LISREL (Jöreskog & Sörbom, 2000; Yuan & Bentler, 1997) and Mplus (Muthén & Muthén, 1998). These programs handle a variety of models such as growth curve models, multilevel models, categorical measures, and missing data as well as measurement models. Approaches to compute power for covariance structure models are described in Kaplan (1995), Jöreskog and Sörbom (2000), Muthén and Curran (1997), and Satorra and Saris (1985).

Several measures of drug use are now available to researchers. Self-report is the method most often used to measure drug use and abuse. Tests from samples of blood, saliva, breath, urine, and hair are important measures of substance use (Harrison & Hughes, 1997). The number of different drugs that can be identified and the accuracy of identification from body samples has increased dramatically (e.g., techniques based on gas chromatography–mass spectrometry [GC-MS] have high specificity and sensitivity for many drugs). The availability of several different measures of drug use should provide more detailed information about the validity of drug use measures. The biological measures cost considerably more than self-report measures and may have considerable ethical concerns that may limit their use. Nevertheless, it is clear that these sophisticated biological approaches to the measurement of drug use will increase in the future, perhaps leading to cheaper more accurate methods.

Most substance abuse research focuses on variables that do not have clear biological measures. These include personality variables (such as rebelliousness, depression, and risk taking), norm variables (such as perceptions of prevalence of drug use), attitude variables (such as positive and negative expectations about drug use), communication variables, and resistance skill variables. For some of these constructs, there is a tradition of careful measurement development and testing such as for measures of depression. For many of the important variables in substance use research, however, there is not a well-developed research literature to support the adequacy of these measures. Measurement of these constructs may be the weakest part of substance abuse research. The problem is exacerbated in substance abuse research where the time allotted to questionnaire administration can be quite small compared to the number of questionnaire items researchers want participants to complete. Often important constructs are measured with three or fewer items or not included at all. One new solution to this problem is to use multiple forms of a survey questionnaire and then use a statistical approach to combine the data from all forms including partially complete data (Graham *et al.*, 1996). More constructs can be measured with multiple forms, although there is considerable data organization required for this approach to be successful.

Another approach to improving the measurement of constructs is the use of behavioral measures in addition to self-report (Palmer *et al.*, 1994). In the measurement of resistance skills

for example, Graham *et al.* (1989) used behavioral observation with several raters in a comprehensive multiple method model for measuring resistance skills.

A final measurement topic is the important distinction between latent and emergent variables (also called scales and indices, respectively) in substance abuse research. As described earlier, latent variables are unobserved constructs that are measured by variables such as individual questionnaire items. As outlined by Widaman and Reise (1997) and Bollen and Lennox (1991), emergent variables refer to variables that are not likely to be predicted by a latent factor but are instead formed by combining items. For example, individual negative life event measures are specified to cause the emergent variable of negative life events, rather than the latent variable of negative life events causing the individual negative life events. The best way to incorporate these measures in covariance structure modeling is not yet clear but some guidance is provided by MacCallum and Browne (1993). This issue is important because there are many variables in substance abuse research that may be more accurately modeled as emergent variables.

MISSING DATA

Often respondents do not complete all questions in a survey. In all longitudinal drug prevention studies, some participants are not measured at follow-up, other participants come in and out of the research design providing measures only periodically (Graham *et al.*, 1997). In a meta-analysis of 84 drug prevention studies, Hansen *et al.* (1990) report average retention rates of 81.4% at 3 months, 73.4% at 1 year, and 67.5% at 3-year follow-up.

Researchers describe missing data in one of three ways leading to different statistical approaches to dealing with missing data (Enders, 2001). Missing completely at random (MCAR) means that the data are missing randomly, for example, random assignment of multiple forms of a questionnaire. If multiple forms are used, then participants will be missing some variables only because they did not receive a certain form. Another type of missing data is Missing at Random (MAR) or accessible missing data, where the data are not missing completely at random but a measure that predicts missingness is in the data. An example of this type of missing data occurs when students graduate from school and are therefore unavailable for measurement only because they have graduated. If the data are not MCAR or MAR, the task is more complicated because the reason for the missing data is unknown or inaccessible, and if the missing data are ignored, the analysis may be inaccurate. Approaches are available for missing data that are nonignorable or not missing at random but they require some knowledge about the potential causes of missingness (Little & Yau, 1996, 1998). Because of the developments in missing data analysis, researchers can now analyze all data, even partially complete data. Furthermore, researchers can conduct more studies with planned missing data. It remains unclear how the methods can handle different amounts of missing data but it is likely that the reason for the missing data (MCAR, MAR, or neither MAR nor MCAR) is more important than the amount of missing data.

Methods that include partial as well as complete data have been used in a variety of substance use studies (Graham *et al.*, 1997; Shafer, 1997). Several software packages are now available including the AMOS general purpose covariance structure program that conducts full information maximum likelihood procedures to adjust for missing data (Arbuckle, 1997). The SAS MIXED procedures can also include partially complete data in many different models. Other programs include PAN, NORM, and MIX, available at Joe Shafer's Pennsylvania State University website, and SOLAS (see Table 1). Some of these procedures, SOLAS for example, will generate files exportable to other computer programs such as SAS and SPSS. Most missing data methods, however, assume that the data are MAR which assumes that a variable explaining the missingness is included in the analysis.

Graham and Donaldson (1993) describe an important missing data analysis methodology when resources are not available to measure all variables from all participants. The data from a small sample with intensive measurement can be combined with a larger sample that includes some but not all the measures in the intensive measurement sample. Statistical power is increased by combining both samples in the analysis. These approaches are ideal for situations where apparatus or the cost of measurement (e.g., drug testing) prohibits measurement of a large sample.

NONNORMAL DATA

Most statistical methods, such as the analysis model described in Eq. 1, assume normally distributed measures. If data that are not normally distributed are analyzed using methods that assume normality, then the standard errors are typically too small and a researcher is more likely to find an effect that is not real, a Type I error (Browne, 1984;Harlow, 1986). Two major procedures are now available to handle measures that are not normally distributed. First, there are statistical approaches based on the values of skewness and kurtosis of variables used in the analysis. Several of these adjustments are available in the EQS (Bentler, 1997) and LISREL (Jöreskog & Sörbom, 2000) computer software packages.

The second approach to dealing with nonnormal data, computer-intensive methods, is growing and holds considerable promise for the analysis of substance abuse research data. In general, these computer-intensive methods use the observed data to determine the significance of an effect and do not make assumptions about underlying distributions (Mooney & Duval, 1993). To illustrate the procedures, assume that a correlation of .2 is found between the number of joints of marijuana smoked in the last month and score on a rebelliousness scale in a sample of 100 adolescents. In a simple bootstrap analysis, a sample of 100 is taken with replacement (which means that the same individual's data may be chosen more than once in one sample) from the original sample of 100. Then a second sample of 100 is taken from the original sample. The process is repeated a large number of times, the correlation is calculated in each bootstrap sample, and the distribution of the bootstrap correlation coefficients is used to determine significance. The 95% confidence limits of the correlation are then the values of the correlation at the 2.5 and 97.5 percentiles in the distribution of bootstrapped correlations (for other bootstrap methods, see Efron, 2000; Efron & Tibshinari, 1993). The bootstrap procedure is included in the AMOS (Arbuckle, 1997), EQS (Bentler, 1997), and LISREL (Jöreskog & Sörbom, 2000) programs. A second group of computer-intensive methods are called randomization tests (Edgington, 1995; Noreen, 1989). Computer-intensive tests such as the bootstrap and randomization tests (Manly, 1998) have not been widely used in substance abuse prevention but seem to be ideally suited to the hypotheses tested and data available in substance abuse prevention research.

COMPREHENSIVE MODELS

Equation 1 includes one dependent variable and three independent variables. Typically models for the development and causes of substance abuse include more variables and propose a more complicated pattern of relationships among these variables. Examples of these more comprehensive models are covariance structure models and growth curve models for the etiology of drug use (Curran *et al.*, 1996; Duncan & Duncan, 1996; Huba & Bentler, 1983). Multivariate growth curve approaches have rarely included more than three growth processes whereas covariance structure models typically include many different latent variables in the same model. Newer computer software no longer limits the number of growth processes, so more complicated growth models should appear in the research literature. One limitation of covariance structure modeling is the potential for misinterpretation of correlational relationships as causal. Given no other information about variables, the causal direction among

variables measured at the same time cannot be known. A related problem is the existence of equivalent models that fit the data equally well but may be quite different from the proposed model (MacCallum *et al.*, 1993; Spirtes *et al.*, 1993). These developments encourage researchers to consider alternative equivalent models in their research.

Probing the Effects of Substance Abuse Prevention

As the field of drug prevention has matured, statistical approaches have been applied to answer detailed questions about how, why, and for whom prevention program effects occur. These methods use the statistical approaches described previously but are adapted to provide detailed information about prevention program effects.

IMPLEMENTATION ANALYSIS

Adequate implementation is critical for the success of any health education curriculum (Kolbe, 1986; Pentz & Trebow, 1991). All of the analyses described previously assume that the program evaluated was implemented properly. In fact, the effect of the program depends on whether the program was implemented as designed (adherence), whether participants received the curriculum (exposure), and whether the program was modified during its implementation (reinvention). Examples of implementation data include the number of sessions attended and the extent to which program deliverers changed the program. Analysis of program effects without measures of implementation may inaccurately suggest that a program failed to have the desired effects.

Several recently developed models incorporate the amount of exposure to an intervention, providing a more fine grained analysis of program exposure. One type of model is the Intention to Treat (ITT) analysis where participants assigned to the treatment group are considered to be in the treated group even though they may not have received any exposure to the program. Other approaches called the LATE (Local Average Treatment Effect) and CACE (Complier Average Causal Effect) attempt to estimate program effects as a function of the amount or type of exposure (Little & Yau, 1998). In this literature, exposure to an intervention is called compliance because of the use of these approaches in the evaluation of drug treatments for disease. In general these approaches use information on implementation as an additional predictor in models such as Eq. 1.

MEDIATION ANALYSIS

Mediating variables are the constructs that a program is designed to change that are hypothesized to cause reductions in substance use (Judd & Kenny, 1981). Hansen (1992) describes mediators typically targeted in substance abuse prevention activities, such as social norms, beliefs about consequences, and resistance skills. For example, prevention programs are designed to engender norms that are less tolerant for drug use and that is hypothesized to reduce drug use. Program effects on substance use are often reported in research papers but program effects on mediators are not often reported. A few studies have tested the entire mediational process by which the program changed drug use. Studies of primary prevention of the use of gateway drugs such as tobacco, alcohol, and marijuana suggest that social norms and beliefs about positive consequences are important mediators of substance abuse prevention programs (Donaldson *et al.*, 1994; MacKinnon *et al.*, 1991). Mediation analysis provides evidence on how the program achieved its effect by testing the hypothesized causal sequence of the program changing the mediator that in turn leads to a change in the outcome (MacKinnon & Dwyer, 1993). Such information increases understanding of the mechanisms that lead to change and determines how prevention programs work so they can be modified to be cost-effective by including only critical components.

Methods to assess mediation for a single mediator (Baron & Kenny, 1986; MacKinnon & Dwyer, 1993) and multiple mediators (Bollen, 1987; MacKinnon, 2000) have been described and these models have been applied in the analysis of drug prevention data (e.g., Botvin, 2000; MacKinnon *et al*., 1991). Experimental approaches to identifying mediating processes by randomly assigning participants to levels of the mediators are described in West and Aiken (1997). There have been a few applications of mediation analysis in growth curve methodology (e.g., Cheong *et al*., 2003). One application of longitudinal growth modeling to mediation analysis examines the relationship between the growth of substance use and the mediator when substance use and the mediating variable are measured repeatedly over time. The growth of substance use and the growth of the mediating variable are considered two concurrent processes that are influenced by the prevention program. The growth of the mediating variable is hypothesized to influence growth of the outcome. In this analysis, the program is modeled to influence the growth of substance use directly and also indirectly via the growth of the mediator. To test the hypothesis, two steps of analysis are conducted (Stoolmiller, 1994). First, a series of univariate longitudinal growth curve models are tested for the mediator and substance use separately. In the second step, the models that adequately describe the growth curve of substance use and the mediator are entered into one multivariate longitudinal growth curve model. Specific hypotheses then test (1) the direct effect of the program on the growth of substance use and the mediator and (2) the mediated effect of the treatment on growth in substance use through growth of the mediator. A potential limitation with this approach is that the slope for the mediator predicts the growth in substance use which implies a concurrent relationship. Another model, requiring additional waves of data, examines the influence of early growth in the mediator on later growth in substance use. For example, change in the mediator from baseline to the first follow-up is hypothesized to predict change in substance use from the first to the second follow-up. More specifically, the treatment effect might change the mediator at an early period and this, in turn, might influence substance use at a later period.

Methods of testing for mediation in single-level models can be successfully adapted for the multilevel case. Multilevel mediation effects are generally similar in magnitude to those generated in single-level mediational analyses, but the standard errors are larger (Krull & MacKinnon, 1999). This results in mediation tests that are appropriately more conservative than single-level tests when there is significant clustering in the data.

MODERATOR ANALYSIS

Equation 1 assumes no interactions, meaning that the effect of the program does not depend on other variables such as age, sex, or rebelliousness. Fortunately, Eq. 1 can be extended to include interaction effects between program exposure, demographic, and other variables. These interactions can be tested in single interaction models and in models adjusted for covariates. Statistical methods to analyze these interactions are described in Aiken and West (1991). Recent methods incorporate latent interaction effects, such as an interaction between program exposure and a latent measure of rebelliousness (Schumacker & Marcoulides, 1998). These methods incorporate the unreliability in the measures comprising the interaction variables to provide a better measure of the interaction by forming variables that represent latent interaction effects.

There have also been considerable developments in methods to assess invariance across subgroups (Millsap, 1995; Widaman & Reise, 1997), which is an important issue when interest lies in the extent to which predictors differ across subgroups such as gender or racial groups. The analyses generally use covariance structure methodology to test the equality or invariance of effects across groups, using a chi-square test comparing the model with and without parameters freed across groups.

Recently, statistical software has been developed to estimate models that include both trajectories of individual change over time and different predictors of change. These models, known as mixture models, represent a new method to examine moderator effects as they allow differential trajectories for subgroups of individuals. The General Growth Mixture Modeling (GGMM; Muthén, 1998) expands conventional growth modeling (Willett & Sayer, 1994), latent class modeling (Clogg, 1995), finite mixture modeling (Laub *et al.*, 1998; Muthén & Shedden, 1999; Nagin *et al.*, 1995; Verbeke & Lesaffre, 1996), and structural equation modeling. The conventional latent growth model assumes individuals come from the same population and follow the same normative growth curve. Although individual differences in the growth curves are captured by the variation in the random growth factors, the conventional growth modeling cannot model heterogenous growth shapes that would indicate qualitatively different development (a moderator effect). In contrast, the GGMM approach allows the modeling of the heterogenous trajectory classes and differential effects of trajectory classes on continuous latent variables. Different trajectory shapes are hypothesized and the probability of each individual's class trajectory membership is calculated as well as effects of background covariates on the different trajectory classes. For these analyses, SAS TRAJ (Jones *et al.*, 1999) and Mplus (Muthén & Muthén, 1998) can be used. The trajectory method has recently been used to identify different trajectories of tobacco use onset (Chassin *et al*., 2000).

Interpreting the Results of Substance Abuse Prevention Research

There have been several trends in the interpretation of substance abuse research that represent refinements of methods introduced earlier. Four related topics are discussed here: the criticisms of hypothesis testing approaches to research and arguments to use confidence limits when interpreting research results, the use of statistical power and effect size to interpret results, the growth of scientific approaches to combine results across studies, and the detailed causal interpretation of research results.

HYPOTHESIS TESTING AND ALTERNATIVES

Recently, there has been criticism of hypothesis testing as the primary goal of research in the social sciences (Harlow *et al.*, 1997; Krantz, 1999; Nickerson, 2000; Wilkinson & the Task Force on Statistical Inference, 1999) and the medical sciences (Bailar & Mosteller, 1988). Hypothesis testing refers to the specification of null (H_0) and alternative hypotheses (H_A) and the use of an inferential statistical test, such as a *t* or *F* statistic, to decide whether the null hypothesis should be rejected or not on the basis of a sample of data. The hypothesis testing approach has been criticized for (1) its fixation on a Type I error rate of .05 (1 out of 20 chance of deciding that an effect is present when it actually is not) as a magical number that indicates whether H_0 should be rejected, (2) its reliance on the binary reject H_0 or not reject H_0 decision, and (3) the influence of sample size on the decision such that tiny effects can be highly statistically significant if sample size is very large. Some researchers advocate the use of confidence limits because the estimate is specified along with some interval to judge the accuracy of the estimate leading to more careful consideration of the magnitude of an effect as well as its statistical significance. It is possible, for example, that the effect size was identical across several studies but the null hypothesis was rejected only a few times. Substance abuse research may be improved by including both confidence limits and hypothesis testing in research publications. A related view suggests that a more reasonable approach may be to compute the probability that a research finding is true using a Bayesian perspective (Malakoff, 1999) and other types of information relevant to the hypothesis. However, the Bayesian approach is criticized for the subjective nature of the prior information used in testing the hypothesis. It is likely that the use of Bayesian approaches will increase in substance abuse prevention as it has in other fields.

STATISTICAL POWER AND EFFECT SIZE

Statistical power, the ability of a research study to detect a real effect, is now part of substance abuse research planning and interpretation. About 40 years ago, Cohen (1962) demonstrated that the power to detect real effects was remarkably low in psychological research, about a 50–50 chance of finding a real effect. Low power was also found in other research areas (Cohen, 1990; Freiman *et al.*, 1978). More recent studies also demonstrate a lack of statistical power to detect effects (Rossi, 1990). In substance use research, Hansen (1992) demonstrated that many drug prevention studies lacked statistical power to detect effects and that, in general, small scale studies miss detecting promising approaches to prevention.

There are three ways that consideration of statistical power will continue to improve substance abuse research. First, in planning a research study, it forces researchers to consider whether the planned sample size and research design are sufficient to find a real effect. Second, it forces researchers to consider how big their effect is rather than just whether the effect reached conventional levels of statistical significance. Third, it provides a manner to evaluate previous studies (e.g., if a study had low statistical power, it is not surprising that a significant effect was not observed).

There are many resources available for power calculations, including Cohen's book (Cohen, 1988) that contains numerous power tables for correlations, analysis of variance, and regression. Statistical software to compute power is also available (Goldstein, 1989; Thomas & Kreb, 1997). A software program called Gpower is now available on the Internet (see Table 1). Widespread knowledge of statistical power in the planning and interpretation of research studies will improve the research conclusions from substance abuse research.

Measures of effect size are used to compare effects across studies that may differ in sample size and the scale of measurement of the dependent variable. Two of the most common effect size measures are the correlation coefficient and the standardized mean difference—the difference between two means divided by the standard error of the difference (Cohen, 1988). The odds ratio and relative risk are measures of effect size for categorical outcome data. Recent studies suggest care should be taken in interpreting effect size in research studies. There are several situations where small effects are important and meaningful. First, small effects in a large population translate into large practical effects. For example, a primary prevention study effect of a 4% difference in new smokers between control and treatment groups translates to thousands of smokers if extended to the entire U.S. population. Therefore, from a public health perspective, small effects can be meaningful. Second, small effects measured by amount of variance explained, or R^2, can actually correspond to important effects especially when categorical variables are analyzed (Abelson, 1985; Cohen, 1988).Rosnow and Rosenthal (1989) noted a research study that demonstrated that aspirin helped prevent heart attacks. The effect of aspirin explained only one third of 1% of the variance, but this translated to 3.4% fewer persons having heart attacks. Examples such as this one suggest that small but statistically significant results should be carefully considered before deeming an effect as too small to be of interest.

META-ANALYSIS

Historically the results of many studies are combined and summarized in a carefully written qualitative review of papers. Meta-analysis is a methodology to quantitatively combine results from many studies to more clearly identify the effects of a phenomenon under study (Bangert-Drowns, 1988). Since Glass (1976) first proposed the use of meta-analysis, there has been a substantial increase in the number of meta-analyses published in refereed journals. Several large meta-analytic studies have been conducted in substance abuse prevention. An important early meta-analysis by Tobler (1986) demonstrated that social-influences-based school

prevention programs led to substantial decreases in substance use when combining across studies, even though not all studies found statistically significant effects (Bangert-Drowns, 1988).A recent NIDA monograph (Bukoski, 1997) on the use of meta-analysis in substance abuse research demonstrates the degree of maturation of this method in prevention research. One early conclusion from meta-analyses is that research publications should include the means in each group, the standard errors, and specifics about the sample, type of prevention program, and its components. When this information is included in a study, subsequent meta-analysis is possible and more convincing. Meta-analysis has been criticized because it often gives equal weight to studies of varying methodological rigor, and the independent variable studied (e.g., the drug prevention program) may vary from study to study. Meta-analysis results can be biased because published studies with significant effects are more likely to be included in meta-analyses than unpublished studies without a significant effect. Another source of bias in meta-analysis is the treatment of effects as fixed when random effects may be more reasonable. Probably the most critical lesson from meta-analysis is that each substance abuse prevention study should report the information (e.g., effect sizes, means, standard deviations) necessary for future meta-analysis.

An important approach to combining results across studies is CSAP's Prevention Enhancement Protocols or PEPS reports. PEPS reports (seeTable 1) are now available on several topics including tobacco prevention, prevention of substance abuse among children and adolescents, and control of alcohol availability (Grover, 1999).A report on media approaches to prevent substance abuse and school-based substance abuse prevention should be finished soon. The purpose of these reports is to translate research results into a form useful for both researchers and practitioners. As part of this work an expert panel of academic researchers and practitioners meets and discusses research findings and place them in one of four levels of evidence on the basis of the research conducted: strong level of evidence, medium level of evidence, suggestive or insufficient evidence, and substantial evidence of ineffectiveness. There are actually several reports for each topic, a reference guide describing research studies and levels of evidence, a practitioner's guide and a parent and community guide. The PEPS reports are important because they represent the critical link between research results and using these results to prevent substance abuse. It is likely that more of these reports will appear and perhaps reports will be updated as more research is conducted.

CAUSAL INTERPRETATION OF EFFECTS

There have been interesting new developments in the detailed causal analysis of research effects. The purpose of these methods is to carefully consider the different information necessary for causal interpretation of results. Many theorists note that causal interpretation is the motivation for research studies even though researchers may not claim that their results provide causal conclusions. One of these models is Rubin's causal model (Rubin, 1974). The model has several basic principles. First, the approach claims that it only makes sense to talk about the effects of cause relative to another cause (e.g., treatment vs. control group). Second, appropriate causal interpretation can only be made when it is possible for individuals to be in either group (e.g., an individual could be in either the program or control group).Third, the cause must temporally precede the effect. The model demonstrates that in most situations, only random assignment can lead to causal interpretation of the effect of a program group compared to a control group. The approach provides a general framework to understand the limitations and strengths of possible causal inferences from any research study (Winship & Morgan, 1999). The approach has been applied to mediation models (Holland, 1988) and to epidemiological measures (Robins & Greenland, 1992). This detailed approach has not yet been applied to substance abuse prevention data but it should help researchers provide context for their research findings based on the type of study conducted.

SUMMARY

The sophistication of the statistical methods applied in substance abuse research is encouraging. Multilevel models allow for the incorporation of effects at different levels such as school and community as well as individual effects. These models allow for some interesting tests of effects across levels such as the effect of community, school, and classroom on individual substance use. It is possible, for example, that the effect of a school norm may be larger than the effect of a classroom norm. If the time until an event occurs is studied in a research project, then survival analysis can be used to understand the predictors of whether and when an event, such as first drug use, has occurred. Similarly, if more than two repeated measures are taken from the participants (or other units such as communities or schools) then trajectories over time can be investigated and prevention effects on these trajectories can be examined. Additional waves of longitudinal data allow for more detailed investigation of trajectories over time. Researchers can hypothesize different trajectories for subgroups of participants such as high-risk groups. New methods allow for the incorporation of categorical dependent variables in comprehensive models for drug use. Mediation analysis clarifies the importance of translating the theoretical basis of a prevention program into the mediating variables that can be measured in a study. If the study design includes measures of the dependent variables (usually drug use) and mediating variables, then mediation analysis can be used to shed light on how a prevention program worked or failed to work.

Future methodologies will likely come from enhancements of the methods described in this paper. A general covariance structure model may be feasible that would include missing data, multiple levels of analysis, different trajectories, implementation measures, mediating processes, and would easily accommodate categorical and continuous variables. The closest models and computer software designed to accomplish this goal are the covariance structure programs of AMOS (Arbuckle, 1997), EQS (Bentler, 1997), LISREL (Jöreskog & Sörbom, 2000), and Mplus (Muthén & Muthén, 1998). Computer intensive methods, such as bootstrapping, are also likely to play a role in future advances in statistical methods.

Improved computing technology will continue to enhance substance abuse prevention research. The Internet is already having an impact on statistical methods in substance abuse prevention. Many websites now include prevention materials. The Internet has the potential of being a very powerful data collection system because it is widely available to a large number of participants. It is likely that future methods research will examine the accuracy of this method of data collection. Several researchers now maintain websites for statistical information and downloading of computer programs, as well as drug prevention information (see Table 1). It is also possible that specific analyses can be run using software at a website so that the researcher submitting the data can learn new statistical methods and the author of the website can gain more exposure to actual prevention research problems. Overall, the availability of e-mail and the Internet along with advancements in computer speed and storage suggest a substantial change in the way substance abuse research is conducted. These changes increase collaboration and communication among researchers making cross-site research easier (see Sambrano *et al.*, 1997, for issues related to cross-site research).

A related future direction in substance abuse research is secondary data analysis. The ease of storage and analysis of large data sets has made it possible for individual researchers to obtain and analyze them. Several national data sets are available and these represent an important resource for substance abuse research (National Center for Health Statistics, 2000). Databases such as the Cochran Controlled Trials Register provide a means to document all research of a certain type (Brown & Liao, 1999; Cochrane Controlled Trials Register, 1999). Similarly, National Institute of Health (NIH, 1995) promotes secondary analysis research studies as a way to extract the maximum amount of information from data that has already been collected.

The purpose of this paper was to provide an overview of some advances in statistical methods for substance abuse prevention. Advances in statistical methods in substance abuse research will continue. The focus on these advanced statistical techniques must be qualified by two considerations. First, randomization of units to conditions will likely remain the best approach to identifying true relationships among variables. Of course, there are many situations where randomization is not possible or unethical but the research is needed (Cook & Campbell, 1979). Nevertheless, the ability to randomly assign units to prevention program exposure will remain a cornerstone of scientific substance abuse research. A second consideration has been called minimally sufficient analysis in a recent report of the American Psychological Association's Task Force on Research Methods (Wilkinson & the Task Force on Statistical Inference, 1999). Minimally sufficient analysis requires that a statistical analysis should be no more complicated than necessary to answer a scientific question. For research results to be understood by a large audience, it is necessary that a widely known statistical technique be used. The statistical methods described in this paper should not entirely replace more straightforward approaches to data analysis. On the other hand, the advances in statistical analysis extract more information and appropriately adjust for the unique aspects of substance abuse prevention research data. In this respect, they may lead to more accurate conclusions from a research study. Nevertheless, the quality of the research question and the idea driving the research are the most important aspects of substance abuse research. Statistical methods are tools to help answer these questions. The tools are getting better.

Acknowledgments

This paper was supported by Substance Abuse and Mental Health Services Administration and the Center for Substance Abuse Prevention (CSAP) as a state of the science review. It was also supported in part by Public Health Service Grant DA09757 from the National Institute on Drug Abuse to Dr David P. MacKinnon.

REFERENCES

Abelson RP. A variance explanation paradox: When a little is a lot. Psychological Bulletin 1985;97:129–133.

Abelson RP, Prentice DA. Contrast tests of interaction hypotheses. Psychological Methods 1997;2:315–328.

Aiken, LS.; West, SG. Multiple regression: Testing and interpreting interactions. Newbury Park, CA: Sage; 1991.

Anderson, ER. Accelerating and maximizing information from short-term longitudinal research. In: Gottman, JM., editor. The analysis of change. Mahwah, NJ: Erlbaum; 1995.

Anselin, L.; Florax, RJGM. New directions in spatial econometrics. Berlin: Springer; 1995.

Arbuckle, JL. AMOS[Computer software]. Chicago: Small Waters; 1997.

Bacik JM, Murphy SA, Anthony JC. Drug use prevention data, missed assessments and survival analysis. Multivariate Behavioral Research 1998;33:573–588.

Bailar JC, Mosteller F. Guidelines for statistical reporting in articles for medical journals: Amplifications and explanations. Annals of Internal Medicine 1988;108:266–273. [PubMed: 3341656]

Bangert-Drowns RL. The effects of school-based substance abuse education: A meta-analysis. Journal of Drug Education 1988;18:243–264. [PubMed: 3058921]

Barcikowski RS. Statistical power with group mean as the unit of analysis. Journal of Educational Statistics 1981;6:267–285.

Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 1986;51:1173–1182. [PubMed: 3806354]

Bentler PM. Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology 1980;31:419–456.

Bentler, PM. EQS for Windows (Version 5.6) [Computer software]. Encino, CA: Multivariate Software, Inc; 1997.

Bentler PM, Newcomb MD. Personality, sexual behavior, and drug use revealed through latent variable methods. Clinical Psychology Review 1986;6:363–385.

Bollen, KA. Total, direct, and indirect effects in structural equation models. In: Clogg, CC., editor. Sociological methodology, 1987. Washington, DC: American Sociological Association; 1987. p. 37-69.

Bollen, KA. Structural equations with latent variables. New York: Wiley; 1989.

Bollen KA. Latent variables in psychology and the social sciences. Annual Review of Psychology 2002;53:605–634.

Bollen KA, Lennox R. Conventional wisdom on measurement: A structural equation perspective. Psychological Bulletin 1991;110:305–314.

Botvin GJ. Preventing drug abuse in schools: Social and competence enhancement approaches targeting individual-level etiologic factors. Addictive Behaviors 2000;25:887–897. [PubMed: 11125777]

Brown CH, Liao J. Principles for designing randomized preventive trials in mental health: An emerging developmental epidemiology paradigm. American Journal of Community Psychology 1999;27:673–710. [PubMed: 10676544]

Brown, H.; Prescott, R. Applied mixed models in medicine. New York: Wiley; 1999.

Brown, CH.; Wang, W.; Muthén, BO.; Dagne, G. Technical Report Prevention Science and Methodology Group. Power calculations on the World Wide Web for longitudinal intervention trials. Department of Epidemiology and Biostatistics, University of South Florida; 2000.

Browne MW. Asymptotically distribution-free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology 1984;37:62–83. [PubMed: 6733054]

Bryk, AS.; Raudenbush, SW. Hierarchical linear models. Applications and data analysis methods. Newbury Park, CA: Sage; 1992.

Bryk, AS.; Raudenbush, SW.; Seltzer, M.; Congdon, RT. An introduction to HLM: Computer program and user's guide. Chicago: University of Chicago; 1988.

Bukoski, W., editor. Meta-analysis of drug abuse prevention programs. NIDA Research Monograph 170. Rockville, MD: U.S. Department of Health and Human Services; 1997. NIH Publication No. 97-4146

Chassin L, Presson CC, Pitts SC, Sherman SJ. The natural history of cigarette smoking from adolescence to adulthood in a Midwestern community sample: Multiple trajectories and their psychosocial correlates. Health Psychology 2000;19:223–231. [PubMed: 10868766]

Cheong J, MacKinnon DP, Khoo S-T. Investigation of mediational processes using parallel process latent growth curve modeling. Structural Equation Modeling 2003;10:238–262. [PubMed: 20157639]

Chou CP, Montgomery S, Pentz MA, Rohrbach LA, Johnson CA, Flay BR, MacKinnon DP. Effects of a community-based prevention program on decreasing drug use in high-risk adolescents. American Journal of Public Health 1998;88:944–948. [PubMed: 9618626]

Clogg, CC. Latent class models. In: Arminger, G.; Clogg, CC.; Sobel, ME., editors. Handbook of statistical modeling for the social and behavioral sciences. New York: Plenum Press; 1995. p. 311-359.

The Cochrane Library. Oxford: Update Software; 1999. Cochrane Controlled Trials Register. (Updated quarterly)

Cohen J. The statistical power of abnormal–social psychological research. Journal of Abnormal and Social Psychology 1962;65:145–153. [PubMed: 13880271]

Cohen, J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum; 1988.

Cohen J. Things I have learned (so far). American Psychologist 1990;45:1304–1312.

Cohen, J.; Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum; 1983.

Collins LM, Wugalter SE. Latent class models for stage-sequential dynamic latent variables. Multivariate Behavioral Research 1992;27:131–157.

Cook, TD.; Campbell, DT. Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally; 1979.

Crocker, L.; Algina, J. Introduction to classical and modern test theory. Orlando, FL: Harcourt, Brace, Jovanovich; 1986.

Curran PJ, Harford TC, Muthén BO. The relation between heavy alcohol use and bar patronage: A latent growth model. Journal of Studies on Alcohol 1996;57:410–418. [PubMed: 8776683]

DiClemente CC, Prochaska JO, Fairhurst SK, Velicer WF, Velasquez MM, Rossi JS. The process of smoking cessation, an analysis of precontemplation, contemplation, and preparation stages of change. Journal of Consulting and Clinical Psychology 1991;59:295–304. [PubMed: 2030191]

Diggle, PJ.; Liang, KY.; Zeger, SL. Analysis of longitudinal data. Oxford: Clarendon Press; 1994.

Donaldson SI, Graham JW, Hanson WB. Testing the generalizability of intervening mechanism theories: Understanding the effects of adolescent drug use prevention interventions. Journal of Behavioral Medicine 1994;17:195–216. [PubMed: 8035452]

Donner A. A regression approach to the analysis of data arising from cluster randomization. International Journal of Epidemiology 1985;14:322–326. [PubMed: 4019000]

Duncan TE, Duncan SC. A multivariate latent growth curve analysis of adolescent substance use. Structural Equation Modeling 1996;3:323–347.

Duncan TE, Duncan SC, Hops H. Analysis of longitudinal data within accelerated longitudinal designs. Psychological Methods 1996;1:236–248.

Duncan, TE.; Duncan, SC.; Strycker, LA.; Li, F.; Alpert, A. An introduction to latent variable growth curve modeling: Concepts, issues, and applications. Mahwah, NJ: Erlbaum; 1999.

Edgington, ES. Randomization tests. New York: Marcel Dekker; 1995.

Efron B. The bootstrap and modern statistics. Journal of the American Statistical Association 2000;95:1293–1296.

Efron, B.; Tibshirani, R. An introduction to the bootstrap. New York: Chapman and Hall; 1993.

EGRET [Computer software]. Seattle: Statistics and Epidemiology Research Corporation. 1991.

Enders CK. A primer on maximum likelihood algorithms available for use with missing data. Structural Equation Modeling 2001;8:128–141.

Freiman JA, Chalmers TC, Smith H, Kuebler RR. The importance of beta, the type 2 error and sample size in the design and interpretation of the randomized control trial. The New England Journal of Medicine 1978;299:690–694. [PubMed: 355881]

Glass GV. Primary, secondary, and meta-analysis of research. Educational Researcher 1976;5:3–8.

Goldstein R. Power and sample size via MS/PC-DOS computers. American Statistician 1989;43:253–260.

Goldstein H. Nonlinear multilevel models with an application to discrete response data. Biometrika 1991;78:45–51.

Goldstein H, Rasbash J. Improved approximations for multilevel models with binary responses. Journal of the Royal Statistical Society, Series A 1996;159:505–513.

Goldstein, H.; Rasbash, J.; Plewis, I.; Draper, D.; Browne, W.; Yang, M.; Woodhouse, G.; Healy, M. A user's guide to MlwiN. London: University of London; 1998.

Graham JW, Donaldson SI. Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data. Journal of Applied Psychology 1993;78:119–128. [PubMed: 8449850]

Graham, JW.; Hofer, SM.; Donaldson, SI.; MacKinnon, DP.; Schafer, JL. Analysis with missing data in prevention research. In: Bryant, KJ.; Windle, M.; West, SG., editors. The science of prevention: Methodological advances from alcohol and substance abuse research. Washington, DC: American Psychological Association; 1997. p. 325-365.

Graham JW, Hofer SM, MacKinnon DP. Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. Multivariate Behavioral Research 1996;31:197–218.

Graham JW, Rohrbach LA, Hansen WB, Flay BR, Johnson CA. Convergent and discriminant validity for assessment of skill in resisting a role play alcohol offer. Behavioral Assessment 1989;11:353–379.

Grover, PL. Preventing problems related to alcohol availability: Environmental approaches. Rockville, MD: Department of Health and Human Services; 1999. (DHHS Publication No. (SMA) 99-3298)

Gruenewald PJ, Millar AB, Treno AJ, Yang Z, Ponicki WR, Roeper P. The geography of availability and driving after drinking. Addiction 1996;91:967–983. [PubMed: 8688823]

Haggard, EA. Intraclass correlation and the analysis of variance. New York: Dryden Press; 1958.

Hansen WB. School-based substance abuse prevention: Review of the state of the art in curriculum, 1980–1990. Health Education Research 1992;7:403–430. [PubMed: 10171672]

Hansen WB, Tobler NS, Graham JW. Attrition in substance abuse prevention research: A meta-analysis of 85 longitudinally followed cohorts. Evaluation Review 1990;14:677–685.

Harlow, LL. Unpublished doctoral dissertation. Los Angeles: University of California; 1986. Behavior of some elliptical theory estimators with nonnormal data in a covariance structures framework: A Monte Carlo study.

Harlow, LL.; Mulaik, SA.; Steiger, JH., editors. What if there were no significance tests?. Mahwah, NJ: Erlbaum; 1997.

Harrison, L.; Hughes, A. The validity of self-reported drug use: Improving the accuracy of survey estimates, NIDA research monograph 167. Rockville, MD: U.S. Department of Health and Human Services; 1997. The validity of self-reported drug use: Improving the accuracy of survey estimates; p. 1-16.DHHS Publication No. 97-4147

Hedeker D, Gibbons RD. MIXREG: A computer program for mixed-effects regression analysis with autocorrelated errors. Computer Methods and Programs in Biomedicine 1996;49:229–252. [PubMed: 8800609]

Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychological Methods 1997;2:64–78.

Holland, PW. Causal inference, path analysis, and recursive structural equations models. In: Clogg, C., editor. Sociological methodology, 1988. Washington, DC: American Sociological Association; 1988. p. 449-484.

Hosmer, DW.; Lemeshow, S. Applied logistic regression analysis. New York: Wiley; 1989.

Hosmer, DW.; Lemeshow, S. Applied survival analysis: Regression modeling of time to event data. New York: Wiley; 1999.

Hsieh FY. Sample size tables for logistic regression. Statistics in Medicine 1989;8:795–802. [PubMed: 2772439]

Huba GJ, Bentler PM. Test of a drug use causal model using asymptotically distribution free methods. Journal of Drug Education 1983;13:3–14.

Hyatt, SL.; Collins, LM. Using latent transition analysis to examine the relationship between perceived parental permissiveness and the onset of substance use. In: Rose, JS.; Chassin, L.; Presson, CC.; Sherman, SJ., editors. Multivariate applications in substance use research. New York: Erlbaum; 2000. p. 141-160.

Jones BL, Nagin DS, Roeder K. A SAS procedure based on mixture models for estimating developmental trajectories. Sociological Methods and Research 2001 May;29:374–393.

Jöreskog, K.; Sörbom, D. LISREL [Computer software]. Lincolnwood, IL: Scientific Software, Inc; 2000.

Judd, CM.; Kenny, DA. Estimating the effects of social interventions. New York: Cambridge University Press; 1981.

Kandel DB. The social demography of drug use. Milbank Quarterly 1991;69:365–414. [PubMed: 1805113]

Kaplan, D. Statistical power in structural equation modeling. In: Hoyle, RH., editor. Structural equation modeling: Concepts, issues, and applications. Newbury Park, CA: Sage; 1995. p. 100-117.

Keppel, G. Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall; 1991.

Kirk, RE. Experimental design: Procedures for the behavioral sciences. Pacific Grove, CA: Brooks/Cole; 1995.

Kolbe LJ. Increasing the impact of school health promotion programs: Emerging research perspectives. Health Education 1986;17:47–52. [PubMed: 3152360]

Kosterman R, Hawkins JD, Guo J, Catalano RF, Abbott RD. The dynamics of alcohol and marijuana initiation: Patterns and predictors of first use in adolescence. American Journal of Public Health 2000;90:360–366. [PubMed: 10705852]

Krantz DH. The null hypothesis testing controversy in psychology. Journal of the American Statistical Association 1999;44:1372–1381.

Krull JL, MacKinnon DP. Multilevel mediation modeling in group-based intervention studies. Evaluation Review 1999;23:418–444. [PubMed: 10558394]

Krull JL, MacKinnon DP. Multilevel modeling of individual and group level mediated effects. Multivariate Behavioral Research 2001;36:249–277.

Laub JH, Nagin DS, Sampson RJ. Trajectories of change in criminal offending: Good marriages and the resistance process. American Sociological Review 1998;63:225–238.

Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD. SAS system for mixed models. Cary, NC: SAS Institute; 1996.

Little, TD.; Schnabel, KU.; Baumert, J. Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples. Mahwah, NJ: Erlbaum; 2000.

Little R, Yau L. Intent-to-treat analysis for longitudinal studies with drop-outs. Biometrics 1996;52:1324–1333. [PubMed: 8962456]

Little RJ, Yau LHY. Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. Psychological Methods 1998;3:147–159.

Lord, FM.; Novick, MR. Statistical theories of mental test scores. Reading, MA: Addison-Wesley; 1968.

MacCallum RC, Browne MW. The use of causal indicators in covariance structure models: Some practical issues. Psychological Bulletin 1993;114:533–541. [PubMed: 8272469]

MacCallum RC, Wegener DT, Uchino BN, Fabrigar LR. The problem of equivalent models in applications of covariance structure analysis. Psychological Bulletin 1993;114:185–199. [PubMed: 8346326]

MacKinnon, DP. Contrasts in multiple mediator models. In: Rose, JS.; Chassin, L.; Presson, CC.; Sherman, SJ., editors. Multivariate applications in substance use research. New York: Erlbaum; 2000. p. 141-160.

MacKinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. Evaluation Review 1993;17:144–158.

MacKinnon DP, Johnson CA, Pentz MA, Dwyer JH, Hansen WB, Flay BR, Wang EY. Mediating mechanisms in a school-based drug prevention program: First-year effects of the Midwestern Prevention Project. Health Psychology 1991;10:164–172. [PubMed: 1879388]

Malakoff D. Bayes offers a new way to make sense of numbers. Science 1999;286:1460–1464. [PubMed: 10610542]

Manly, BFJ. Randomization, bootstrap and Monte Carlo methods in biology. London: Chapman & Hall; 1998.

Millsap RE. Measurement invariance, predictive invariance, and the duality paradox. Multivariate Behavioral Research 1995;30:577–605.

Mooney, CZ.; Duval, RD. Bootstrapping: A nonparametric approach to statistical inference. Newbury Park, CA: Sage; 1993.

Murray, DM. Design and analysis of group-randomized trials. New York: Oxford University Press; 1998.

Murray DM, Rooney BL, Hannan PJ, Peterson AV, Ary DV, Biglan A, Botvin GJ, Evans RI, Flay BR, Futterman R, Getz JG, Marek PM, Orlandi M, Pentz MA, Perry CL, Schinke SP. Intraclass correlation among common measures of adolescent smoking: Estimates, correlates, and applications in smoking prevention studies. American Journal of Epidemiology 1994;140:1038–1050. [PubMed: 7985652]

Muthén, BO. Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. Paper presented at the conference of New Methods for the Analysis of Change, Pennsylvania State University; 1998, October; PA.

Muthén BO, Curran PJ. General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. Psychological Methods 1997;2:371–402.

Muthén BO, Khoo S-T. Longitudinal studies of achievement growth using latent variable modeling. Learning and Individual Differences 1998;10:73–101.

Muthén BO, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics 1999;55:463–469. [PubMed: 11318201]

Muthén, LK.; Muthén, BO. MPlus [Computer software]. Mooresville, IN: Scientific Software; 1998.

Nagin DS, Farrington D, Moffit T. Life-course trajectories of different types of offenders. Criminology 1995;33:111–139.

National Center for Health Statistics. Current legislative authorities of the national center for health statistics. Hyattsville, MD: U.S. Department of Health and Human Services, Center for Disease Control and Prevention; 2000.

Newcomb, MD.; Bentler, PM. Consequences of adolescent drug use: Impact on the lives of young adults. Newbury Park, CA: Sage; 1988.

Nickerson RS. Null hypothesis significance testing: A review of an old and continuing controversy. Psychological Methods 2000;5:241–301. [PubMed: 10937333]

NIH Program Announcement. Secondary analysis of alcohol abuse prevention research data. Rockville, MD: 1995. NIH Guide, PAR-95-024

Noreen, EW. Computer-intensive methods for testing hypotheses: An introduction. New York: John Wiley; 1989.

Palmer RF, Dwyer JH, Semmer N. A measurement model of adolescent smoking. Addictive Behaviors 1994;19:477–489. [PubMed: 7832006]

Palmer RF, Graham JW, White EL, Hansen WB. Applying multilevel analytic strategies in adolescent substance use prevention research. Preventive Medicine 1998;27:328–336. [PubMed: 9612823]

Pentz, MA.; Trebow, E. Implementation issues in drug abuse prevention research. In: Leukefeld, CG.; Bukoski, WJ., editors. Drug abuse prevention intervention research: Methodological issues, NIDA Research Monograph 107. Washington, DC: U.S. Department of Health and Human Services; 1991. p. 123-139.DHHS Publication No. 91-1761

Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology 1992;3:143–155. [PubMed: 1576220]

Rogosa, D. Myths about longitudinal research. In: Schaie, KW.; Campbell, RT.; Meredith, W.; Rawlings, SC., editors. Methodological issues in aging research. New York: Springer; 1988. p. 171-209.

Rosnow RL, Rosenthal R. Statistical procedures and the justification of knowledge in psychological science. American Psychologist 1989;44:1276–1284.

Rossi JS. Statistical power of psychological research: What have we gained in 20 years? Journal of Consulting and Clinical Psychology 1990;58:646–656. [PubMed: 2254513]

Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 1974;66:688–701.

Sambrano, S. Center for Substance Abuse Prevention national cross-site evaluation of high risk youth programs: Research design. St. Louis, MO: EMT Associates; 1996.

Sambrano S, Springer JF, Hermann J. Informing the next generation of prevention programs: CSAP'S cross-site evaluation of the 1994–1995 high risk youth grantees. Journal of Community Psychology 1997;25:375–395.

SAS (Version 8.0) [Computer software]. Cary, NC: SAS Institute, Inc; 1999.

SAS Institute, Inc. SAS® technical report P-229, SAS/STAT® Software: Changes and enhancements, release 6.07. Cary, NC: SAS Institute, Inc; 1992. The MIXED procedure. p. 287-366.

SAS Institute, Inc. SAS/STAT® Software: Changes and enhancements through release 6.12. Cary, NC: SAS Institute, Inc; 1997. The GENMOD Procedure; p. 247-348.

Satorra A, Saris WE. Power of the likelihood ratio test in covariance structure analysis. Psychometrika 1985;50:83–90.

Schumacker, RE.; Marcoulides, GA. Interaction and nonlinear effects in structural equation modeling. Mahwah, NJ: Erlbaum; 1998.

Selvin, S. Statistical analysis of epidemiologic data. Oxford: Oxford University Press; 1996.

Shafer, JL. Analysis of incomplete multivariate data. New York: Chapman & Hall/CRC Press; 1997.

Siddiqui O, Flay BR, Hu FB. Factors affecting attrition in longitudinal smoking prevention study. Preventive Medicine 1996;25(5):554–560. [PubMed: 8888323]

Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. Journal of Educational and Behavioral Statistics 1998;23:323–355.

Spirtes, P.; Glymour, C.; Scheines, R. Causation, prediction, and search. New York: Springer-Verlag; 1993.

SPSS (Version 9.0) [Computer software]. Chicago: SPSS, Inc; 1999.

Stoolmiller M. Antisocial behavior, delinquent peer association, and unsupervised wandering for boys: Growth and change from childhood to early adolescence. Multivariate Behavioral Research 1994;29:263–288.

Thomas L, Kreb CJ. A review of statistical power analysis software. Bulletin of the Ecological Society of America 1997;78:126–139.

Tobler NS. Meta-analysis of 143 adolescent drug prevention programs: Quantitative outcome results of program participants compared to a control or comparison group. Journal of Drug Issues 1986;16:537–567.

Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association 1996;91:217–221.

Voas RB, Marques PR, Tippetts AS, Beirness DJ. The Alberta interlock program: The evaluation of a province-wide program on DUI recidivism. Addiction 1999;94:1857–1867.

West, SG.; Aiken, LS. Toward understanding individual effects in multicomponent prevention programs: Design and analysis strategies. In: Bryant, KJ.; Windle, M.; West, SG., editors. The science of prevention: Methodological advances from alcohol and substance abuse research. Washington, DC: American Psychological Association; 1997. p. 167-209.

Widaman, KF.; Reise, SP. Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In: Bryant, KJ.; Windle, M.; West, SG., editors. The science of prevention: Methodological advances from alcohol and substance abuse research. Washington, DC: American Psychological Association; 1997. p. 281-324.

Wilkinson L. the Task Force on Statistical Inference. Statistical methods in psychology journals: Guidelines and explanations. American Psychologist 1999;54:594–604.

Willett JB, Sayer AG. Using covariance structure analysis to detect correlates and predictors of individual change over time. Psychological Bulletin 1994;116:363–381.

Willett JB, Singer JD, Martin NC. The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. Development and Psychopathology 1998;10:395–426. [PubMed: 9635230]

Winship C, Morgan SL. The estimation of causal effects from observational data. Annual Review of Sociology 1999;25:659–707.

Yin RK, Kaftarian SJ. Introduction: Challenges of community-based program outcome evaluations. Evaluation and Program Planning 1997;20 293-197.

Yin RK, Kaftarian SJ, Yu P, Jansen MA. Outcomes from CSAP's community partnership program: Findings from the national cross-site evaluation. Evaluation and Program Planning 1997;20:345–355.

Yuan KH, Bentler PM. Mean and covariance structure analysis: Theoretical and practical improvements. Journal of the American Statistical Association 1997;92:767–774.

Table 1

Internet Sites With Statistical Information

Topic	URL	Software
General statistics sites		
Prevention Science and Methodology Group	http://www.psmg.hsc.usf.edu	
David MacKinnon Prevention Stat Class	http://www.public.asu.edu/~davidpm/classes/psy536/536syl.htm	
Royal Holloway Psychology Department	http://www.pc.rhbnc.ac.uk/jt/stathome.html	
Michael Friendly's Statistics Links	http://www.math.yorku.ca/SCS/friendly.html	
Clay Helberg's Statistics Links	http://www.execpc.com/~helberg/statistics.html	
David C. Howell's Links	http://www.uvm.edu/~dhowell/StatPages/Archives.html	
Polytomous regression – John Hendrickx	http://www.xs4all.nl/~jhckx	X
SAS	http://www.sas.com	X
SPSS	http://www.spss.com	X
Multilevel modeling/clustering		
Don Hedeker's Website	http://tigger.uic.edu/~hedeker	
MIXOR and MIXREG	http://tigger.uic/edu/~hedeker/mix.html	X
BUGS Project	http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml	X
HLM	http://www.ssicentral.com/hlm/hlm.htm	X
MLwin	http://multilevel.ioe.ac.uk/index.html	X
Geographical Clustering	http://www.atsdr.cdc.gov/HS/cluster.html	X
Longitudinal models		
MPlus	http://www.statmodel.com/	X
Proc Traj	http://www.stat.cmu.edu/~bjones/traj.html	X
Survival analysis		
Statistics at Square One	http://www.bmj.com/statsbk/12.html	
StatSoft Electronic Textbook	http://www.statsoft.com/textbook/stsurvan.html	
Measurement error		
LISREL	http://www.ssicentral.com/lisrel/mainlis.htm	X
EQS	http://www.mvsoft.com/	X
AMOS	http://www.spss.com/amos	X
Mx–Virginia Commonwealth University	http://views.vcu.edu/mx/index.html	X
Tips on survey design	http://www.surveysystem.com/sdesign.htm	
Missing data		
SOLAS	http://www.statsolusa.com/solas/solas.htm	X
Penn State Methodology Center	http://methcenter.psu.edu/mde.html	X
Joe Schafer at Penn State	http://www.stat.psu.edu/~jls/misoftwa.html	X
Nonnormal data		
Resampling methods	http://www.resample.com	X
Implementation analysis		
Intention to Treat Primer	http://www.childrens-mercy.org/stats/ask/intention.asp	
Mediation analysis		
Research in Prevention Laboratory	http://www.public.asu.edu/~davidpm/ripl/mediate.htm	

Topic	URL	Software
David Kenny's Mediation Page	http://nw3.nai.net/~dakenny/mediate.htm	
Meta-analysis		
Study Database Analyzer	http://www.assess.com/Software/Meta-Analysis.htm	X
Comprehensive Meta-Analysis	http://www.meta-analysis.com	X
Prevention Enhancement Protocols	http://text.nlm.nih.gov/ftrs/dbaccess/csap	
Confidence limits		
STATISTICA's online textbook	http://www.statsoft.com/textbook/stpowan.html	
Jim Steiger's Homepage	http://www.interchg.ubc.ca/steiger/homepage.htm	X
Statistical power		
Gpower	http://www.psycho.uni-duesseldorf.de/aap/projects/gpower	X
DSS Research Sample Size Calculators	http://www.dssresearch.com/SampleSize/default.asp	X
Power analysis	http://www.PowerAndPrecision.com	X