International Journal of Database Theory and Application
Vol. 11, No. 4 (2018), pp.1-12
http://dx.doi.org/10.14257/ijdta.2018.11.4.01

# Database Theories as Frameworks for Statistical Data Analysis of Research on Alcohol Use Among Adolescents

Ngozi M. Eze[1], Eucharia N. Aye[2], Nkechi Egenti[2], Amaka B. Ezeanwu[1], Nkiru Ohia[3], Ngozi H. Chinweuba[2], Nneka Charity Nwosu[2], Okechukwu O. Nwaubani[4*], Bernedeth N. Ezegbe[4], Kennedy O. Ololo[5], Chukwudozie J. Emeka[6], Kay C.N. Onyechi[2] and Bartholomew C. Nwefuru[2]

*[1]Department of Home Economics and Hospitality Management Education, University of Nigeria, Nsukka*
*[2]Department of Educational Foundations, University of Nigeria, Nsukka*
*[3]Institute of African Studies, University of Nigeria, Nsukka*
*[4]Department of Social Science Education, University of Nigeria, Nsukka*
*[5]Department of Sociology/Psychology/Criminology & Security Studies, Federal University Ndufu Alike Ikwo, Ebonyi State, Nigeria*
*[6]Community Secondary School, Nando, Anambra State.*
**okechukwu.nwaubani@unn.edu.ng*

***Abstract***

*This paper reviewed the relational model and dependence theory of database management in the context of statistical data analysis of alcohol use research. The relational model of database management promotes an understanding of the means to manage data via a structure and language in accordance with first-order logic. Another theory of database management known as dependence theory helps to ascertain the occurrence of a phenomenon. The paper focused on statistical data analysis for studies on alcohol use among adolescents with anchorage on these database theories. Statistical data interpretation with good statistical tools such as intraclass, longitudinal measurement, multilevel, covariance structure model among others can help in providing reliable information about adolescents' abuse of alcohol. Longitudinal assessment of alcohol abuse is essential for the investigation of the advancement of alcohol abuse among adolescents because there is a lot of procedural improvement in the interpretation of longitudinal data and various scholars and researchers have demonstrated the uses of longitudinal data as a successful means of identifying preventive measures for limiting alcohol intake by adolescents. Thus, wider knowledge of database theories will boost the planning, analyzing, interpreting and evaluating of alcohol studies which will in turn help to standardize research conclusion about adolescents' alcohol intake.*

***Keywords****: Database theory, Dependence theory of database management, Theory of relational model of database management, Statistical data analysis, Alcohol use, Adolescents, Linear and logistic regression models*

## 1. Introduction

### 1.1. Relational Model and Dependence Theories of Database Management

Statistical data analysis procedures and studies on adolescents' alcohol use can be viewed and understood with good knowledge of database theories. One of the database

Received (June 16, 2018), Review Result (August 22, 2018), Accepted (August 31, 2018)
* Corresponding Author

ISSN: 2005-4270 IJDTA

theories is the relational model of database management which, according to Armstrong is a model for managing data via a structure and language in accordance with first-order logic. The model was first theorized by E. F. Codd at IBM in 1970 as a general model of data. The theorist propounded relational database theory with three assumption or principles of getting hidden information. These elements are called three-valued logic; true, false, missing/null. However, the theorist in 1990 later increased the principle to four elements as true, false, missing but applicable and missing but inapplicable. The major purpose of creating these principles is to ensure that missing information by means of a structure and language which is in line with first-order predicate logic will have a flow order [1].

The tenets of relational database model are that the representation of every single data is by mathematical n-ary relations and an n-ary relation as a subset of the Cartesian product of n domains. Mathematically, thoughts related to such data are revealed in double-valued predicate logic. Thus, there are two potential valued judgments for each assumption which are, either true or false. According to this database model, data are curated through a relational algebra or relational calculus, these being equivalent in expressive power. Also, this model notes that similar data records are merged collectively by means of a key. The aim is to offer a declarative strategy for the specification of data and queries. That is, users straightforwardly affirm what information the database holds and what information they require from it, and allow the database management system software to be in charge of the description of data structures for storage of the data as well as retrieval procedures for answering queries. Some relational database utilizes an SQL data definition and query language. These systems execute an engineering approximation to the relational model. A table in an SQL database schema corresponds to a predicate variable; the contents of a table to a relation; key constraints, other constraints, and SQL queries correspond to predicates. However, SQL databases digress from the relational model in many ways [2].

The relational model of the database allows the creation of a reliable and rational representation of information. Reliability is attained by including declared constraints in the database design, a process known as a logical schema. This theory proposes a process of database normalization wherein a design with specific sought-after characteristics could be chosen from a set of rationally corresponding options. The access plans and other implementation and operation information are taken care of by the DBMS engine and are not observed in the rational model [1, 2]. This is different from the usual practice for SQL DBMSs whereby performance tuning does need modification to the logical model. The fundamental relational building block is the domain or unprocessed information type. Relational database according to Beeri, Fagin, Howard [3] ensures the stability of a relational database is enforced, not by rules built into the requests that utilize it, but rather by constraints, incorporated as part of the logical schema and enforced by the DBMS for all applications. A relational database is also interpreted as a representation of the extension of some predicate as a set of true propositions that can be made by restoring each free variable in that predicate by a name because the model is also called extension. The relational database model is pertinent for use based on the law of excluded middle which stipulated that anything that is not true is false and anything not false is true[4].

Another theory of database management known as dependence theory helps to ascertain the occurrence of a phenomenon. The theory of database dependencies started with the introduction of functional dependencies by Codd in 1972 [5]. Dependency theory focuses on the database implication and optimization problems related to logical constraints, commonly called dependencies. Such dependencies include functional dependencies, which form the foundation of keys to database relations. Another important form of dependencies is the multi-valued dependencies. Thus, dependency database theory focuses on generalization of super keys[5].

 

## 2. Statistical Models and Statistical Instruments

### 2.1. Linear and Logistic Regression Models

Linear and logistic regression model are statistical tools that can be used to ascertain prevalence cases of alcohol use and therefore can be used in understanding issues about excessive use of alcohol among adolescents [6]. The difficulty of focusing on the individual or substances abused and getting reliable information from different statistical tools has remained a concern for an increase in adolescents' alcohol use. Researchers can apply covariance structure modelling to qualitatively boost ways of preventing adolescents' substance abuse. It has also been suggested that in reducing reporters' bias in proffering remedy to alcohol abuse, reporters can use biochemical pointers and covariance structure models for understandability of results [6].

Logistic regression analysis is another statistical data analysis for the study of alcohol use among adolescent, though, it is complicated as a result of modelling difficulty associated with repeated observation. The approach for repeated modelling may need generalized estimate equation. Alcohol abuse preventive measure is learned through the GEE model which appropriate models dependencies in repeated categorical unprocessed information. Currently, the GEE model is included in the SAS programme GENMOD making the method very much faster to carry out. A developed a method called latent transition analysis that generates stage-logical models of adolescents' specifically increase unprocessed information. The approach is an extension of Guttman scaling. Guttman scaling is widely applied to test whether drug use grows from drugs such as tobacco and alcohol to harder drugs such as heroin and cocaine. Latent transition analysis tests hypothesize about the ordering of drug use initiation as well as predictors of the ordering. This approach seems ideal to test theories that postulate these unique stages in the increase from addiction to non-compulsion [6].

### 2.2. Considering Statistical Instruments for Observing Alcohol Research Outcome

A World Health Organization (WHO) study showed that 11% of the student had drunk excessively or gotten drunk at least twice; about 22% of the adolescents interviewed reported having been drunk at least once [7]. Another survey by WHO did show that 18% of 15-year-old adolescents had already tried marijuana during some period of life. Usually, reports on drug use among students show that drugs like alcohol, marijuana and tobacco are mostly used by males, while females use amphetamines and anxiolytics more often. The overuse of alcohol by adolescents is the reason behind investigating a thorough means by which statistical data analysis can help in preventing alcohol abuse among adolescents. According to Yin and Kaftarian[8], adolescents are mostly found in the church, school, clinics, communities and other social places because of their dependency nature and thereby can be identified in the group using multilevel statistical data application.

The quality and efficiency of statistical instruments for observing outcome is restrictive [9]. It is restrictive as Yi j is the dependent measure for the ith individual at dimension j, β1 codes the consequence of the baseline assessment of the dependent variable (X1), β2 codes the result of avoidance programme (X2), β3 codes the result of a covariate such as age (X3), and ei j is the remaining error for the ith individual at the jth measurement. The approximation of alcohol abuse programme result is equivalent to β2 and a t-test of the implication of the programme result is obtained by dividing the approximate of β2 by its predictable standard error. The model assumes randomization of adolescents to the situation, autonomous observations, dependable and correct procedures, and whole unprocessed information. To Abelson and Prentice [9], if there are few or many measurements after baseline, the model can be used to determine programme results at any stage of assessment adjusted for the baseline assessment. Another version of this model includes multiple dependent variables on the left-hand side of the equation where the additional variables correspond to additional stages of weighing. Specifically, the $\beta 2$ coefficient for the program effect coded in X2 can be increased to include both linear and quadratic structures to model the strength or weakness in the programme result over time. Similarly, the results of booster programming can be deal with difference. The statistical data procedures and analyzing of the model is taught in most graduate research programmes and is elicited in previous articles [10]. Many statistical programming packages such as the SAS and SPSS are measures of making analysis more easy and accurate result. The statistical capacity of calculations to determine the number of respondents needed in a research study to asses programme results of a sample size is also explained in the works and software programme available to compute power for these designs [11]. Thus, a frequent dependent variable in alcohol abuse investigation is specifically on the area of dichotomous variable coding whether or not a respondent indulges in the consumption of alcohol. Cohen [11] asserted that one of the most common means of obtaining reliable information on alcohol and other substance preventative measure is logistic regression. Cohen [11] maintained that studies on substance abuse preventive measures may have positive result, especially when tested with logistic regression. On the contrary, Hosmer and Lemeshow [12] noted that logistic regression statistical data is problematic to apply because it is only taught in higher institutions and requires adding several computer programmes to calculate strength for logistic regression analysis.

### 2.3. Intraclass Correlations and Multilevel Analyses

Understanding the level of dependency among adolescents may be through the application of intraclass correlation. Murray et al. [13] stated that constructive intraclass correlation results in rejecting falsely a true null hypothesis. Positive intraclass is likely to fall in inflated rates of Type I error for regression and logistic statistical structure [14]. Even if the intraclass correlation becomes visibly little, it still has the capacity to induce Type I errors. Thus, going by challenge face with the application of intraclass correlation considering the clustering level of adolescents in the school and other social places, interpretation of adolescents' alcohol abuse may be erroneous leading to the wrong conclusion that a programme has a positive effect. Yin and Kaftarian [8] reported the application of intraclass correlation in the discovery of alcohol abuse preventive measures in a cross-site assessment of CSAP's high-risk adolescent programmes and CSAP's communal joint venture. The scholars noted that many levels of unprocessed information can be processed procedural using multilevel analysis. The statistical tool is as well used in the case of individuals in schools (Level 1) and school (Level 2) levels.

According to Yin and Kaftarian [8] at stage one, a linear model is specified for adolescents within a given learning environment. Trying to ascertain due procedure to be taken in preventing adolescents from alcohol abuse, the statistical model in structural form are taken to be random and differ as a role of predictors at the learning environment capacity. Thus, this equations sum up these relations: adolescents Level 1:$Yij = \beta 0j + \beta 1jXij + eij$=(2)-School Level 2:$\beta 0j = \gamma 00 + \gamma 01Wj + u0j$=(3)=$\beta 1j = \gamma 10 + u1j$=(4) then, subscript i refers to adolescents and subscript j refers to learning environment. It explained how adolescents-level predictors in the X matrix (like age, gender, or other covariates) and school-level predictors in the W matrix (like assessment to programme or manage groups or other school attributes) can be interpreting concurrently in a multilevel analysis. The estimation of error terms at both levels of the model (ei j at the individual level and u0 j and u1 j at the social environment level) permits for a nonzero intraclass correlation to be integrated into the interpretation [8].

 

Besides bending for a nonzero intraclass correlation, multilevel models can be used to examine effects at the different levels (the social environment effect on individuals). Classes for learning can be included as a third level in the interpretation permitting for the assessment of school and class learning effects on individuals. To add, these essential cross-level effects have not been tackled in a comprehensive way in the alcohol abuse prevention [6]. The possibility for these cross-level associations makes it essential that scholars or researchers include measures of the learning environment, a class for learning interaction, and any other possible clustering variable in their exercise. The hypotheses about different levels also apply to the clustering of different adolescents in learning environment, homes, or hospitals. Power calculations that bend for the effects of gathering are explained for both continuous and specific dependent variables. Brown et al. [15] stated multilevel data interpretation helps to calculate power that could be adjusted for the effects of clustering in adolescents' alcohol misuse preventive measures.

There are numerous programming software that carries out multilevel interpretation. Programmes especially designed for the interpretation of multilevel models include HLM and MLn [16]. The authors added that two programmes contain subroutines that accommodate categorical outcome variables. Multilevel model is a special intervention case of the general mixed model, the MIXED procedure of the SAS software method (SAS Institute, Inc.) can be applied to produce the required estimates. The SAS MIXED software can be added with the SAS GLMMIX subroutine to conduct multilevel interpretation of a specific dependent variable. More so, multilevel analysis give wider information regarding how adolescents' alcohol abuse can be prevented through geographical analysis where relevant messages or information on the venue of data collection is associated with interpretation on geographic analyses of the relationship between the alcohol establishment density and alcohol-related problems reported that multilevel analysis is effective interpretative measure to ascertain how adolescents can be restricted from consumption of alcohol regardless of location involved [16].

Longitudinal assessment of alcohol abuse is essential for the investigation of the development of alcohol abuse. Jöreskog and Sörbom [17] were of the view that another way to understand how alcohol abuse can be prevented is through the use of latent growth curve modelling (LGM) in the structural equation modelling framework using LISREL. Note that LGM is a statistical method to measure, elucidate, and describe adolescents' dissimilarities in change over time. When data are collected on a good number of adolescents over many observations, LGM assesses adolescents' growth or development curves. The growth model contains two levels of assessment, the repeated observations or within-individual model (Level 1) and the individual-level or between-individual model (Level 2). At Level 1 all members of the population are assumed to have routes of a collective form but each person can have different values for the growth parameters that include initial status and rate of change. The growth parameters then become the result variables at Level 2, where they are predicted by individual-level attributes. The categorization of the growth curve in terms of levels of analysis demonstrates its relationship to multilevel modelling. In fact, multilevel modelling and covariance structure modelling are used to analyze growth curve models. Other latent growth curve methods permit the proper analysis of data that includes repeated measures from many cohorts of respondents across many waves. This method includes a level of analysis for the cohort as well as the within-individual and between-individual levels [18].

## 3. Structural Models, Measured Variables and Latent Variables in Alcohol Research

Two types of covariance variable structure modelling exist as measured variables and latent variables. Crocker and Algina [19] stated that when measures are not reliable, observed relationships among them may be significantly varied and that measures that are not valid lead to the wrong conclusion about relationships. They reported that an advanced approach for alcohol abuse involves an application of covariance structure. Latent variables are abstract, theoretical constructs that cannot be directly measured but must be inferred on the basis of concrete variables. Measured variables are experimental elements, such as questionnaire items. A measurement model is applied to standardize more true measures of latent variables by categorizing how measured elements are related to the latent elements. When many indicators are applied to measure a latent variable, the latent variable is more dependable than each character item. The structural model specifies the relationships among these latent factors. Covariance structure models have been applied widely in alcohol abuse research. There has been an understandable progression in the statistical software to calculate these models [19]. Appropriate statistical data analysis computer software programmes for understanding substance abuse outcomes are LISREL and Mplus [17]. LISREL and Mplus statistical programmes manage different models like growth curve models, multilevel models, categorical measures, and missing data as well as measurement models [19].

Harrison and Hughes [20] noted that tests from samples of blood, saliva, breath, urine, and hair are essential measures of alcohol use. The accessibility of many different measures of alcohol use should provide more comprehensive information about the validity of alcohol use measures. The biological measures charge significantly more than personal-report measures and may have significant ethical worries that may limit their applicability. However, it is understandable that applying these biological methods to the measurement of alcohol use will increase in the future, possibly leading to the cheaper more accurate approach of substance abuse prevention [20]. To better understand alcohol abuse, Graham et al. [21] stated that researches should focus more on biological measures. Graham et al. [21] further opined that the measurement of some constructs (non-biological measures) may be the weakest part of alcohol abuse research. The challenge is intensified in alcohol abuse research where the time given to questionnaire administration can be quite small compared to the number of cluster items researchers want respondents to complete. Always essential constructs are measured with three or fewer items or not included at all. A recent headway to this problem is to use multiple forms of a survey questionnaire and then use a statistical method to combine the data from all forms including partially complete data [21].

Another method to strengthen the measurement of constructs is the use of behavioural measures. Graham et al. [21] applied behavioural observation in the measurement of resistance skills with many raters in a comprehensive multiple approach model for measuring resistance skills. There is an essential distinction between latent and emergent variables. Latent variables are abstract constructs which are measured by other variables like individual questionnaire items. Emergent variables as variables are not likely to be predicted by a latent factor but are instead formed by combining items. To MacCallum and Browne[22], the harmonious means to assemble these measures in covariance structure modelling is not yet clear but some direction may be provided through accurate modelled as emergent variables.

## 4. Taking Care of Missing Statistical Data in Alcohol Research

Responding to the over changing measure in ascertaining adequate means of curbing alcohol abuse through statistical data analysis, Graham et al. [21] noted that people who fill survey questionnaires most times neglect research procedural instruction on completion of all questions. Graham *et al.*, [21] stated that in some cases like in longitudinal drug prevention researches, participants may not be measured at follow-up. In the same light, Hansen et al. [23] reported in a meta-analysis of 84 drug preventive types of research that average retention rates of 81.4% at 3 months, 73.4% at 1 year, and 67.5% at 3-year follow-up. Thus, missing data can lead to wrong interpretation and conclusion regarding preventive measures of alcohol abuse. The author restated that missing completely at random (MCAR) suggest that the data are losing randomly, for instance, random assessment of multiple forms of a questionnaire. Thus, when multiple forms are applied, then respondents will lose some variables because they avoided certain form. There is also another form of missing data known as Missing at Random (MAR) or locatable missing data where the data does not lose wholly at random but a measure that predicts missingness is in the unprocessed information. For instance, this form of missing unprocessed information occurs when adolescents' student from school graduates and is therefore out of school for measurement only because they have graduated. If the data are not MCAR or MAR, the task is more complex because the reason for the missing data is unidentified or hard to find, and if the missing data are neglected, the analysis may be erroneous. Approaches are available for missing data that are not neglectable or not missing at random but they require some knowledge about the potential causes of missingness. In addressing the issue of missing data in order to enhance quality result on preventive measures on alcohol abuse among adolescents, many scholars believes that many software packages are now accessible including the AMOS general purpose covariance structure programme that conducts full information maximum likelihood procedures to adjust for missing data. Graham et al. [21] opined that approaches that include partial as well as full data have been used differently in studies of alcohol abuse. The SAS MIXED methodology can as well include partial or whole data in several varieties of models. Other programmes include PAN, NORM, MIX, and SOLAS which can generate files exportable to other computer programs such as SAS and SPSS. Most missing data methods, nevertheless, suppose that the data are MAR which presumes that a variable elucidating the missingness is included in the interpretation. Graham and Donaldson [24] explained a significant missing data analysis procedure when possessions are not accessible to measure all variables from all respondents. For these authors, the data from a little sample with rigorous measurement can be joined with a bigger sample that includes some but not all the measures in the rigorous measurement sample. For alcohol abuse research among adolescents, controllable measures can be availed through the articulated statistical data analysis.

According to Harlow [25] many statistical procedures like analysis model presume that data are rightly disseminated but if data are not rightly dispersed, and can be analyzed using approaches that presume rightfully; if on the contrary, then the standard errors are typically too little and a researcher is more likely to find an effect that is not true, a Type I error. Two main approaches are now available to handle measures that are not rightly dispersed. First, there are statistical procedures based on the ideals of skewness and kurtosis and LISREL[17]. The second approach to dealing with non-right data, which is computer-intensive approach is growing and holds substantial promise for the analysis of alcohol abuse data. In general, computer-intensive method uses the observed data to determine the importance of an effect and do not make assumptions about underlying distributions. To illustrate the approach, presume that a correlation of .3 is found between the number of combines 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 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 estimated in each bootstrap sample, and the allocation 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 allocation of bootstrapped correlations. The bootstrap procedure is included in the AMOS, EQS and LISREL programs [17].

MacCallum *et al.*, [26] noted that data analysis of alcohol abuse research among adolescents demands comprehensive models of multivariate growth curve procedures which hardly ever included more than three increase procedures whereas covariance structure models typically include several varieties of latent variables in the same model. MacCallum *et al.*, [26] further pointed out that one restriction of covariance structure modelling is the possibility for misinterpretation of correlational relationships as causal and given no other information about variables, the causal direction among variables measured at the same time cannot be disclosed. Latest computer software no longer restricts the number of growth processes.

## 5. Considering Mediators in the Statistical Analysis of Alcohol Research Data

In alcohol research, mediating variables are essential in the analysis of constructs as a programme designed to change the hypothesis to cause reductions in alcohol use. Hansen [27] pointed out that mediators classically aimed at alcohol abuse prevention activities include social norms, assumptions regarding effects, and resistance skills. For example, preventive programmes are designed to bring about norms that are less broadminded for drug use and that is hypothesized to limit drug use. Programme effects on drug use are often reported in investigative research work but programme effects on mediators are not usually reported. A few studies have tested the entire meditational process by which the programme changed drug use. Studies of primary preventive measures of alcohol use initiation suggest that social norms and beliefs about positive consequences are essential mediators of alcohol abuse preventive programmes. Mediation analysis gives proof on how the programme achieved its results by testing the hypothesized causal sequence of the programme changing the mediator that in turn leads to a change in the result [6]. Such information boosts knowledge of the key variables that lead to change and determines how preventive programmes work so that they can be modified to be cost-effective by including only critical components.

It has been noted that mediation and multiple mediators works efficiently in the area of profiling prevention measures to reduce drug abuse [6]. Experts reported that in experimental methods, observing mediators by means of randomly assigning respondents to grades of the mediators, that there has been a few applicability of mediation interpretation in an increase curve procedure. Corroborating, Cheong et al. [28] stated that in growth curve methodology, little application of mediation analysis has been noticed. The authors stressed further that when alcohol consumption and mediating variables are measured several times over time, that longitudinal growth modelling can be applied to mediation analysis in order to assess the relationship between the growth of substance use and the mediator. Stoolmiller [29] reported that the increase in alcohol consumption and the rise of the mediating variables are seen as simultaneous processes that are influenced by the preventive programme measures. The author further opined that the growth or increase of mediating variable is presupposed to influence the growth of outcome because the programme is modelled to influence the growth of substance consumption both directly and indirectly. Stoolmiller [29] stated that to test hypothesis on mediation regarding alcohol and substance abuse of any kind, two main steps of analysis are required – series of univariate longitudinal growth curve models are tested for the mediator and substance use separately. In the second step, the models which efficiently explain the growth curve of substance consumption and the mediator are input into one multivariate longitudinal growth curve model. Another special hypothesis then examines the following programmes: the direct effect of the programme on the increase or growth of substance consumption and the mediator, and mediated effect of the treatment on increase or growth in alcohol consumption through the increase of the mediator.

Stoolmiller [29] stressing further noted that a potential restriction with this method is that the slope for the mediator predicts the increase in alcohol consumption which suggest a simultaneous relationship because another model may require extra waves of data, investigate the effects of an early increase in the mediator on later increase in alcohol consumption. For instance, change in the mediator from baseline to the first follow-up is guessed to predict change in alcohol use from the first to the second follow-up. More specifically, the treatment influence might change the mediator at an early time and this, in turn, might affect substance use at a later time. According to Krull and MacKinnon [6] approaches of examining or testing for mediation in single-level models can be accomplished by adopting the multilevel case. Krull and MacKinnon added that multilevel mediation influences are generally alike in magnitude to those generated in single-level mediational analyses with a larger standard error which expressed that mediation tests/examinations are appropriately more conservative that single-level examinations when there is vital data gathered [6]. Mediation analysis tries to bring out interaction effects by bringing in latent interaction effects, like the interaction between programme may show no dependence on other variables such as age, sex, or rebelliousness. The interaction incorporates the unpredictability in the measures comprising the interaction variables to provide a better measure of the interaction by forming variables that replace latent interaction effects [6].

Widaman and Reise[30] asserted that there has been a favourably development in ways to assess and understand invariance across many groups which is essential when interest lies in the extent to which predictors differ across different groups such as gender, racial groups among others since the analysis generally use covariance structure methodology to examine the equality or invariance of effects across different groups with chi-square test comparing the model with and without parameters freed across different groups. The general growth mixture modeling (GGMM) is recently developed to estimate models that include both trajectories of individual change over time and different predictors of change and to represent a new approach to examine moderator effects as they allow differential trajectories for different groups of elements or individuals which expands conventional growth modeling, latent class modeling, finite mixture modeling and structural equation modeling. The conventional latent growth model presumes that adolescents come from a common group or population and follow common normative increase curve[30].

## 6. Considering Statistical Power and Effect Sizes in Alcohol Research

Statistical power defines research strength to detect true effects which would work as part of substance abuse research planning and interpretation. Cohen [11] established that the strength to detect true effects was remarkably small in psychological work, adding that about a 50–50 chance of finding a true effect. The author remarked that there is a low statistical strength to detect effects. In drug consumption investigation, Hansen [27] opined that numerous drug prevention studies lacked the statistical strength to detect effects and that little scale studies miss detecting lasting methodology or approaches to drug prevention. While considering statistical strength, three major ways can only improve substance abuse investigation which includes; making sure that planned sample and research design are appropriate for a research study, allowing researchers to consider how large their effects rather than whether the effects reached a conventional level of statistical significance, and giving chance to evaluate previous studies [27]. Thus, wider knowledge of statistical strength in the planning, analyzing, interpreting and evaluating studies will help standardize research conclusions in alcohol abuse investigation.

Cohen [11] stated that effects size is the measure used to compare effects in different studies that might differ in sample size and measurement of the scale of the dependent variable while the most common two effects size measures are the correlation coefficient and the mean standardized difference. Cohen added that the odd ratio and relative risk are measures of effect size for a specific outcome. Abelson and Prentice [9] noted that interpreting effect size in research studies demands carefulness in order to limit error because some situation might make small effects relevant like small effects in a big population translate into big practical effects. For instance, a fundamental control investigative effect of a 4% difference in new alcohol consumers between control and treatment groups translates to many alcohol consumers if expanded to a given population. Thereby, from a general health understanding, little effects can be meaningful and other effects are on a measure of small effects by the amount of variance can actually correspond to vital effects especially when categorical variables are analyzed.

## 7. Study Implications for Alcohol Use Research

Statistical procedures in data analysis include adequate planning, designing, collecting data, analyzing, drawing significant interpretation and reporting of the research findings. The statistical analysis offers understanding and deep sense to the conflicting ideas or senseless numbers, thereby giving hope to a hopeless or lifeless data. The effects and conclusion are simple only if proper statistical tests are applied or initiated. The basic fact about statistical data analysis concerning a preventive measure to alcohol abuse is to skillfully understand different statistical tools before its applicability for generation of data and its analysis. There are many modern and improved statistical procedures currently in use for ascertaining alcohol use trend, and other substance abuse preventive measures, though, some assumed quality statistical data interpretation are sometimes delimited on revealing alcohol and other substance abuse preventive measures. To address alcohol abuse, there should be constant procedures for analyzing alcohol data from adolescents.

If adolescents have sufficient knowledge about the negative effects of alcohol consumption, they may not like to associate themselves with irresponsible decision-making among peers concerning consumption of alcohol. It is imperatively necessary to create conducive learning that would integrate alcohol preventive measures, it is also necessary to understand the phenomenon, which should be included in any health education intervention programmes. The government, non-governmental agencies and traditional rulers even town union agencies should work together in campaigning against adolescents' alcohol intake. The adolescents informal and formal knowledge of alcohol abuse needs a balance information so as to adolescents should implement necessary information on how to say no to alcohol abuse will be informed.

Besides other information on alcohol preventive measure amongst adolescent, there is also need give more detailed information or insight into alcohol prevention policies and programmes aimed at influencing alcohol consumption amongst adolescents. For this rationale, intraclass, multilevel among other data analyses of adolescents has been put into consideration which to an extent differs with analyses of effective policies and programmes in Europe using multilevel governance statistical data analysis. The data analysis of multilevel platform is the sharing of policymaking competences in a system of negotiation between nested governments at different levels including supranational, national, regional and local and private actors like nongovernmental organizations, producers, consumers, citizens amongst other relevant bodies. Multilevel governance is also relevant in another sense, as in this new paradigm of multilevel governance, horizontal governance arrangements gain weight and civil society organizations become more important. Many environmental strategies which prevent adolescent alcohol abuse have been developed in collaboration with civil society, social partners, nongovernmental organizations and other relevant organizations. Local and national governments are only active in setting up the preconditions for providing information about the prevention of alcohol abuse, or by supporting specific groups. All relevant bodies should campaign against alcohol.

From scholarly viewpoints, it seems like multilevel statistical data analysis will allow researchers to bring several countries together in order to investigate how alcohol abuse amongst adolescents can be prevented. In this respect, multilevel policy analysis can be carried out by analysing the policies, programmes and interventions used towards the prevention of alcohol and other substance abuse with questions such as which national policies do national governments pursue with regard to adolescents' alcohol consumption?, which programmes and interventions target the different risk factors (in families, schools and communities)? which programmes and interventions target the adolescents' behaviours, and which programmes and interventions are effective at preventing underage drinking?. Thus, using multilevel statistical data analysis would help in supplying answers to the above questions, such response will determine suggestions, recommendation and conclusion on how adolescents alcohol abuse can be prevented. Finally, opinion from scholars suggests that there is a need to investigate how statistical data can provide more harmonious ways of curbing alcohol and drug abuse among adolescents which is the strong brain behind this work.

## 8. Conclusion

This paper reviewed the relational model and dependence theory of database management in the context of statistical data analysis of alcohol abuse research. Considering the available literature from the reviewed works, the innovation in the world of statistical data analysis concerning alcohol abuse among adolescents is quite impressing. The paper identified an intraclass correlation, mediation modelling, regression, multilevel models, general growth mixture modelling among many others as statistical data analyses tools for adolescents' alcohol abuse studies. Longitudinal assessment of alcohol abuse is essential for the investigation of the development of alcohol abuse because there a lot of procedural development in the interpretation of longitudinal data and many scholars and researchers have demonstrated the uses of longitudinal data as a successful means of identifying preventive measures for limiting alcohol intake by adolescents with the aids of some database theories such as relational and dependency database theory.

## References

[1] W. W. Armstrong, "Dependency Structures of Data Base Relationships", In Proceedings IFIP Congress 74, Stockholm, Sweden **(1974)**, pp. 580–583.

[2] S. J. Rissanen, "Independent Components of Relations", ACM Transactions on Database Systems, vol. 2, no. 4 **(1977)**, pp. 317–325.

[3] C. Beeri, R. Fagin and J. Howard, "A complete axiomatization for functional and multivalued dependencies in database relations. Proceedings of the ACM SIGMOD Conference, **(1977),** pp. 47-61

[4] S. K. M. Wong, C. J. Butz, and Y. Xiang. Automated database schema design using mined data dependencies. J. Amer. Soc. Inform. Sci., vol. 49 **(1998)**, pp. 455–470.

[5] E. F. Codd, "Relational Completeness of Data Base Sublanguages", Data Base Systems, Courant Computer Symposia Series, Prentice-Hall, Englewood Cliffs, New Jersey, Vol. 6, **(1972),** pp. 65-98.

[6] J. L. Krull, and D. P. MacKinnon, "Multilevel" modelling of individual and group level mediated effects", Multivariate Behavioral Research, vol. 36 **(2001)**, pp. 249–277.

[7] World Health Organization, "Inequalities in young people´s health. Health Behavior in School-Aged Children", International Report from 2005-2006. Health Policy for Children and Adolescents, no. 5 **(2008)**.

[8] R. K. Yin and S. J. Kaftarian, "Introduction" Challenges of community-based program outcome evaluations", Evaluation and Program Planning. vol. 20 **(1997)**, pp. 293-197.

[9] R. P. Abelson, and D. A. Prentice, "Contrast tests of interaction hypotheses", Psychological Methods, vol. 2 **(1997)**, pp. 315–328.

[10] J. Cohen, and P. Cohen, "Applied" multiple regression/correlation analysis for the behavioural sciences", Hillsdale, NJ: Erlbaum **(1983)**.
[11] J. Cohen, "Statistical power analysis for the behavioral sciences", Hillsdale, NJ: Lawrence Earlbaum Associates **(1988)**, 2nd ed.
[12] D. W. Hosmer, and S. Lemeshow, "Applied logistic regression analysis", New York: Wiley; **(1989)**.
[13] D. M. Murray, B. L. Rooney, P. J. Hannan, A. V. Peterson, D. V. Ary, A. Biglan, G. J. Botvin, R. I. Evans, B. R. Flay, R. Futterman, J. G. Getz, P. M. Marek, M. Orlandi, M. A. Pentz, C. L. Perry and S. P. Schinke, "Intraclass correlation among common measures of adolescent smoking: Estimates, correlates, and applications in smoking prevention studies", American Journal of Epidemiology, vol. 140 **(1994)**, pp. 1038–1050.
[14] R. S. Barcikowski, "Statistical power with group mean as the unit of analysis", Journal of Educational Statistics, vol. 6 **(1981)**, pp. 267–285.
[15] C. H. Brown, W. Wang, B. O. Muthén and G. Dagne, "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)**.
[16] A. S. Bryk, S. W. Raudenbush, M. Seltzer and R. T. Congdon, "An introduction to HLM: Computer program and user's guide", Chicago: University of Chicago; **(1988)**.
[17] K. Jöreskog, and D. Sörbom, "LISREL Computer software" Lincolnwood, IL: Scientific Software, Inc; **(2000)**.
[18] A. S. Bryk and S. W. Raudenbush, "Hierarchical linear models. Applications and data analysis methods. Newbury Park, CA: Sage; **(1992)**.
[19] L. Crocker and J. Algina, "Introduction to classical and modern test theory". Orlando, FL: Harcourt, Brace, Jovanovich; **(1986)**.
[20] L. Harrison, and A. Hughes, "The validity of self-reported drug use: Improving the accuracy of survey estimates", NIDA Research Monograph 167, no. 97-4147, **(1997)**, pp. 1–16.
[21] J. W. Graham, S. M. Hofer and D. P. MacKinnon, "Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures", Multivariate Behavioral Research vol. 31 **(1996)**, pp. 197–218.
[22] R. C. MacCallum and M. W. Browne, "The use of causal indicators in covariance structure models: Some practical issues", Psychological Bulletin, vol. 114 **(1993)**, pp. 533–541.
[23] W. B. Hansen, N. S. Tobler and J. W. Graham, "Attrition" in substance abuse prevention research: A meta-analysis of 85 longitudinally followed cohorts. Evaluation Review, vol. 14 **(1990)**, pp. 677–685.
[24] J. W. Graham, and S. I. Donaldson, "Evaluating" interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data", Journal of Applied Psychology, vol. 78 **(1993)**, pp. 119–128.
[25] L. L. Harlow, "Behavior of some elliptical theory estimators with nonnormal data in a covariance structures framework: A Monte Carlo study", Unpublished doctoral dissertation. Los Angeles: University of California; **(1986)**.
[26] R. C. MacCallum, D. T. Wegener, B. N. Uchino, and L. R. Fabrigar, "The problem of equivalent models in applications of covariance structure analysis", Psychological Bulletin, 14 **(1993)**, pp. 185–199.
[27] W.B. Hansen. School-based substance abuse prevention: Review of the state of the art in curriculum, 1980–1990. Health Education Research, vol. 7 **(1992)**, pp. 403–430.
[28] J. Cheong, D. P. MacKinnon, and S-T. Khoo, "Investigation of mediational processes using parallel process latent growth curve modelling", Structural Equation Modeling, vol. 10 **(2003)**, pp. 238–262.
[29] M. Stoolmiller, "Antisocial behaviour, delinquent peer association, and unsupervised wandering for boys: Growth and change from childhood to early adolescence", Multivariate Behavioral Research, vol. 29 **(1994),** pp.263–288.
[30] K. F. Widaman, and R. P. Reise, "Exploring the measurement invariance of psychological instruments: Applications in the substance use domain", Edited Bryant KJ, Windle M, West SG, Washington, DC: American Psychological Association; **(1997)**, pp. 281–324.

 