Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Curr HIV Res. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

Curr HIV Res. 2016 ; 14(2): 85–92.

Analytic Strategies to Evaluate the Association of Time-Varying Exposures to HIV-Related Outcomes: Alcohol Consumption as an Example

Robert L. Cook[*,1], **Natalie E. Kelso**[1], **Babette A. Brumback**[2], and **Xinguang Chen**[1]

[1]Department of Epidemiology, College of Public Health and Health Professions and College of Medicine, University of Florida, Gainesville, FL, USA

[2]Department of Biostatistics, College of Public Health and Health Professions and College of Medicine, University of Florida, Gainesville, FL, USA

Background

As persons with HIV are living longer, there is a growing need to investigate factors associated with chronic disease, rate of disease progression and survivorship. Many risk factors for this high-risk population change over time, such as participation in treatment, alcohol consumption and drug abuse. Longitudinal datasets are increasingly available, particularly clinical data that contain multiple observations of health exposures and outcomes over time. Several analytic options are available for assessment of longitudinal data; however, it can be challenging to choose the appropriate analytic method for specific combinations of research questions and types of data. The purpose of this review is to help researchers choose the appropriate methods to analyze longitudinal data, using alcohol consumption as an example of a time-varying exposure variable. When selecting the optimal analytic method, one must consider aspects of exposure (e.g. timing, pattern, and amount) and outcome (fixed or time-varying), while also addressing minimizing bias. In this article, we will describe several analytic approaches for longitudinal data, including developmental trajectory analysis, generalized estimating equations, and mixed effect models. For each analytic strategy, we describe appropriate situations to use the method and provide an example that demonstrates the use of the method. Clinical data related to alcohol consumption and HIV are used to illustrate these methods.

Keywords

Alcohol; generalized estimating equations; generalized linear mixed models; group-based trajectories; HIV; longitudinal; marginal structural models; time-varying exposure

[*]Address correspondence to this author at the Department of Epidemiology, University of Florida, 2004 Mowry Road, P.O. Box: 100231, Gainesville, FL USA; Tel/Fax: 1+ 352 273 5869, 1+ 352 273 5365; cookrl@.ufl.edu.

CONFLICT OF INTEREST

The authors Cook RL, Kelso NE, Brumback BA and Chen X report no real or perceived vested interests that relate to this article that could be construed as a conflict of interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

1. INTRODUCTION

Persons living with HIV (PLWH) are living longer [1, 2]. As they age, PLWH are at risk for a range of clinical events, including incidence of specific conditions (*e.g.*, cardiovascular disease, liver disease, and dementia) and intermediate measures of clinical outcomes (*e.g.*, CD4 count, viral load, blood pressure, depressive symptoms, neurocognitive function). Studies are needed to identify the effect of a range of treatments and interventions, as well as other risk exposures on health outcomes. Many types of exposures cannot be evaluated by traditional randomized trials; they may change over time, and could occur at different time-points across the lifespan from adolescence, young adulthood, and into older age. Behavioral risk factors, such as alcohol consumption, represent exposures that change over time and that influence a range of clinical outcomes [3].

A growing number of datasets that contain real-world observational data among PLWH are available. Typical examples include the Multicenter AIDS Cohort Study, Women's Interagency HIV Study, and Veterans Aging Cohort Study cohorts [4-6]. With datasets such as these, researchers can obtain information regarding changes in exposure and outcome to test a range of study hypotheses. However, it can be challenging to choose the appropriate analytic method for specific combinations of research questions and types of data. Datasets most amenable to the study of changing exposures feature multiple waves of data, clear distinction of time between measures, and repeated assessments of the exposure of interest at the different time-points [7].

The purpose of this review is to overview several analytic methods that may be considered for longitudinal analysis of exposures that change over time, to discuss strengths and limitations of each method, and to provide examples of their use. The purpose is to help researchers choose the appropriate method to analyze longitudinal data. When selecting the optimal analytic method, researchers must consider aspects of exposure (*e.g.* timing, pattern, and amount), and outcome (fixed or time-varying), while also addressing different types of confounding and minimizing bias. In Section 2 of this article, we will first overview some key issues related to exposure, outcome, and bias that may influence one's choice of analytic strategy. In Section 3, several analytic approaches to longitudinal data are described, including specific strengths and limitations of each method. To better describe these methods, we use alcohol consumption as an example of a time-varying exposure. Because alcohol consumption varies over time, measurement of this exposure is complex. Alcohol consumption is highly prevalent among PLWH [8] and has been associated with biological and behavioral influences on health outcomes related to HIV/AIDS [9].

2. KEY ISSUES IN ANALYTIC STRATEGY

2.1. Exposure Measures

Exposures or interventions could be events that occur once or that could occur over different periods of time. Over time, the exposure could be more intense during some stages of life, and completely absent during others. From a causal inference perspective, it is critical that the exposure occurs before the presence of the outcome, and this can sometimes be challenging within longitudinal datasets. To select appropriate methods to analyze such data,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

three key issues related to the exposure measure must be addressed: quantity, timing, and pattern [7]. Here, we will use alcohol exposure as an example to discuss these three issues.

2.1.1. Quantity of Exposure—Commonly-used alcohol exposure measurements seek to define the amount (*e.g.* number of drinks, grams of alcohol) that an individual consumes either within a specific time period (*e.g.* month, week, day) or an amount of exposure over a longer period of time (*e.g.* lifetime). To date, the majority of research on alcohol exposure as a risk factor for long-term health outcomes has used an "amount"-based assessment at a single time-point. For example, recent systematic reviews on the association of alcohol and cardiovascular disease outcomes in PLWH [10], and the general population [11] found that nearly every study on the topic relied on an alcohol measurement that was obtained at a single time-point, with most studies using dichotomous categories of "drinkers" and "non-drinkers". Similar limitations are noted in a recent review of the relationship between alcohol consumption and colorectal cancer in the general population [12], and the relationships between alcohol consumption and HIV-related outcomes such as time to ART initiation and survival [13], and rate of CD4 decline over time [8].

Measuring alcohol consumption at a single time-point and assuming that this single measure represents the best "exposure" amount for that individual is likely biased, compared to a measure that incorporates exposure over a longer time period. For example, a person could have consumed a heavy amount of alcohol between age 20 and 40, and then quit due to health reasons. A measure of alcohol taken at age 45 would classify this person as a "non-drinker", which could bias the results of a study of a long-term outcome (*e.g.* cancer incidence) if the previous drinking history is not taken into account [3, 14].

Longitudinal assessment of exposure may provide a more accurate reflection of the true exposure over time. A cumulative amount of exposure over a longer time period can be measured at a single assessment, based on recall of the subject, or can be created by summing exposure data from multiple time-points. For example, the Lifetime Drinking History [15] is a structured interview where the subject recalls patterns of alcohol consumption from the first year of regular drinking to the present. Some investigators have sought to assess the relationship between a cumulative measure of alcohol exposure and risk of liver fibrosis in PLWH [16, 17]. Other investigators have used longitudinal data with repeated measures to create a lifetime cumulative amount. For example, investigators used data from the Nurse's Health Study to create a cumulative average of alcohol intake by averaging alcohol use over time beginning in 1980 [18]. The authors concluded that the cumulative average, *versus* a single baseline assessment measure, provided the most linear and consistent associations with the outcome, and provided more statistical power by using assessments throughout all follow-up periods [18]. While using a cumulative measure of exposure is likely more accurate than a measure at a single time-point, the cumulative measure may not account for variations in exposure that occur within people.

An important consideration related to the "amount" of exposure is whether the association of the exposure to the outcome is linear (each unit of exposure increase causes an increased risk of outcome) or non-linear. For example, alcohol consumption could be associated with a reduced risk (or no risk) of an outcome when the use is consistently low in dose, but

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

associated with a much greater risk when used in high doses. Some analytic strategies assume that there is a linear association between the amount of exposure (*e.g.* drinks per week) and the outcome of interest (*e.g.* HIV disease progression). If the "true" association of the exposure to the outcome of interest is not linear, some types of analyses could present biased results. Alternatively, alcohol exposure may be treated as a categorical variable, with increasing levels of exposure (*e.g.* none, low-moderate, heavy, etc.).

2.1.2. Timing of Exposure—Timing of an exposure reflects when an exposure occurred, such as age or specific period (e.g. the year of 1990) at first exposure, and is essential to determining the causal inference between exposure and outcome. The temporal association of the exposure (alcohol consumption) with the outcome of interest is an important consideration, as most causal models require that an exposure precede a specific outcome. Measuring both exposure and outcome at multiple time-points helps to ensure that the exposure precedes the outcome. However, in many situations, the outcome may occur first and this leads to a change in the exposure. For example, a person may develop liver cirrhosis and then stop drinking because of the illness (often referred to as a "sick-quitter"). An alcohol assessment done at a subsequent visit might classify this person as a non-drinker, and thus conclude that "non-drinkers" have a high risk for liver cirrhosis. Temporal associations may be very short (or even simultaneous) if the outcome of interest is also time-varying (*e.g.* condom use on a specific night in association with the amount of alcohol consumed), or very long (*e.g.* the association of binge drinking in adolescence to incidence of breast cancer as an adult).

The impact of an exposure on a health outcome could vary according to the age of an individual [14]. For example, alcohol consumption could have a very different effect on long-term brain outcomes when exposure occurs during adolescence than if the exposure occurred later in life. One can assess this relative impact of alcohol consumption at different periods of life. For example, the Nurse's Health Study asked participants about the usual number of alcoholic drinks per week prior to age 40, and then compared the risk of breast cancer based on exposure prior to age 40 and exposure after age 40 [18]. The difference in time between assessments could also impact the choice of analytic strategies, as alcohol measurements could be taken anywhere from hourly (or more frequent) to only once or twice over a lifetime. Finally, in some situations the exposure could vary as a function of calendar year, or month. For example, drinking behavior could be different in the summer compared to the rest of the year, or drinking may vary according to social norms within specific periods of time.

2.1.3. Pattern of Exposure—Pattern of exposure refers to a collective assessment that incorporates both amount and timing for a given individual. Exposure patterns can take into account whether one increases or decreases drinking over time, and whether one consumes larger or smaller amounts over time. To capture patterns, the timing of multiple assessments will depend on how often the exposure changes over time. Patterns of drinking can be measured through multiple time-points that range from daily to semi-annually; yearly assessments are less reliable to depict patterns of alcohol consumption over time. One can create a single summary measure for a given individual that attempts to assign a specific

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

pattern of drinking over time to an individual (*e.g.* a trajectory analysis). Some of the more common trajectory patterns could include increasing, declining, persistent high and persistent low drinking levels. However, drinking behavior is complex over time, and other patterns should also be considered, such as waved patterns or off and on patterns [19].

As opposed to summing or graphing the repeated measurements to describe cumulative alcohol consumption, investigators can allow alcohol consumption at each time-point to be described independently, while controlling for autocorrelation that occurs with repeated measures. These types of analyses (*e.g.* generalized estimating equations, mixed effects models) allow one to see individual changes in alcohol consumption from baseline to follow-up time-points that would have been otherwise lost if using a cumulative measure of alcohol exposure. When using repeated measures of alcohol use, it is most beneficial to have 3 or more follow-ups measurements in addition to baseline, with shorter periods between visits to accurately assess the potential of linear, quadratic, and cubic changes that occur over time.

2.2. Outcome Measures

The majority of outcomes assessed in longitudinal data can be categorized as binary or continuous. Binary outcomes are typically events that occur or do not occur and that do not change over time, such as incidence of myocardial infarction or cancer. Continuous outcomes are typically time varying and are indicative of a pathological process, such as depressive symptoms or HIV RNA viral load. Several issues should be considered when analyzing longitudinal data, including bias, confounding, and missing data of the outcome of interest.

2.2.1. Bias—Observational data are subject to a range of potential biases, and some analytic strategies are better than others at bias mitigation. Using longitudinal data rather than cross sectional is one way to mitigate measurement error, for example. Confounding, selection, or measurement bias may occur when standard regression approaches are used to estimate effects of time-varying exposures [3]. Many of the biases, such as measurement bias, or selection bias, are difficult to avoid when using datasets that are already corrected. Investigators should recognize the strengths and limitations of measures that exist in datasets, and be careful when comparing populations that have different exposures when persons are not randomized to the exposure of interest.

2.2.2. Confounding—Confounding is a type of bias in observational studies, and failing to properly adjust for confounding could lead to incorrect inferences about the magnitude and direction of an exposure-outcome association. Confounding variables can be either measured (with existing data present in a dataset), or unmeasured (an individual or social domain for which there are no data in the dataset). Confounding can occur at a point in time or it can be time-dependent. When a confounder is time-dependent, it is simultaneously a confounder and an intermediate variable on the pathway from a previous exposure to a future outcome. For example, marginal structural models are one approach to observational data analysis that is designed to handle time dependent confounding.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2.2.3. Missing Data—By far the most significant limitation to prospective data collection is the large potential for missing data. Missing data are often differentiated into categories of "missing completely at random," "missing at random" and "missing not at random". Missing data are assumed missing completely at random when no variable in the dataset can predict who will have a missing value; missing at random occurs when missing values in the dataset are correlated with some other variable(s) measured and available in the dataset. Missing not at random occurs when missing values in the dataset can be predicted by variable(s) that are not in the dataset. Data are hardly ever missing completely at random, and therefore results will be biased in analytic models that require this to be the case. Dealing with dropout (loss to follow-up) also varies. Some analytic strategies assume that the slope (trajectory) of repeated outcomes prior to dropout continues after dropout, whereas other analyses only use complete data within the dataset.

3. OVERVIEW OF ANALYTIC OPTIONS

Several different types of analytic approaches to analyzing longitudinal data now exist, according to whether the exposure variable and the outcome variable are fixed (time invariant), or time-variant (Table 1). The appropriate analytic strategy should also consider whether the exposure and the outcome variables are normally distributed within a population. Generalized Linear Models (GLM) were designed to address data with non-normal outcome distributions [20]. The standard GLM assumes that the observations are uncorrelated; however, many longitudinal datasets include repeated measures within an individual. Two examples of extensions of GLM to address correlated data are GEE and GLMM.

3.1. Generalized Estimating Equations

Generalized estimating equations (GEE) is a method of estimating parameters of a marginal (population averaged) model and extends generalized linear modeling (GLM) by specifying correlations between repeated measures within subjects [21]. Most GEE analytic programs require the researcher to choose from one of four types of correlations (independent, unstructured, exchangeable, or autoregressive) [22]. Choosing the wrong correlation structure could result in inflated estimates of standard error, although the fixed effect is typically consistent across all four types of correlations [21]. One limitation of GEE is that it will only produce fixed effects, or population averaged parameters. However, GEE has strengths such that subjects do not need to have the same number of observations or have the same follow-up time-points and data can be missing at random (MAR) for unbiased estimates [23]. The Quasi-likelihood under the Independence model Criterion (QIC) statistic is either built into GEE analysis or available as a macro in most data analysis packages, and can be used to assess goodness of fit and compare nested models [24].

As an example of GEE, Sullivan and colleagues [25] aimed to assess the effect of different categories of alcohol consumption on depressive symptoms in patients with and without HIV overtime. Data were collected at four time-points (baseline, one-, two-, and three-year follow-ups) between 2002 and 2008 through questionnaires and administrative and clinical records. Alcohol consumption was defined using several criteria including alcohol abuse and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

dependence (extracted from clinical records using ICD-9 codes), hazardous drinking defined as an AUDIT score of at least 5 (females) or 7 (males); and binge drinking (6 or more drinks in one occasion, 3 or more times a year). Depressive symptoms were defined as a having a PHQ-9 score of greater than 9. Using repeated measures of depressive symptoms as a binary outcome, the investigators assessed alcohol use as a fixed effect (time-invariant) on the time-varying outcome of depressive symptoms. An unstructured covariance matrix for the repeated measures of depressive symptoms was specified. This study indicated that, on average, changes in alcohol consumption categories were positively associated with changes in depressive symptoms over time [25]. See Kahler *et al.* [26], Moriya *et al.* [27], Tsui *et al.* [28], and Seth *et al.* [29] for other relevant examples of GEE.

3.2. Generalized Linear Mixed Models

While GEE models analyze the effect of covariates on average across the population, generalized linear mixed models (GLMM) are used to estimate the effect of an exposure within an individual as well as across a population. GLMMs are an extension of GLMs that incorporate both fixed and random effects. Typical effects, such as in GLMs, are fixed, but an effect may also be random, in that the individual effects in some persons may differ from the average population effects. GLMM allows researchers to explore variability of individual effects within a population of interest by specifying random effects. While the dependent variable must be measured longitudinally, it can be continuous or binary. Similar to GEE, there are no biases related to missing data if the data are missing at random, subjects do not need to have the same number of observations or have the same follow-up time-points, and goodness of fit statistics between nested models can be compared. Also, repeated measure correlation structures for the predictors and outcome can be specified, as previously mentioned in the discussion of GEE. When using GLMM, random effects are assumed to be independent of fixed effect covariates and normally distributed with constant variance [7]. Singer & Willett [7] provide descriptive steps for assessing whether data meet these assumptions. See Finucane, Samet, and Horton [30] for a more detailed overview of GLMM, specifically in HIV and alcohol research.

As an example of GLMM, Kowalski *et al.* [31] sought to study the longitudinal association between alcohol consumption and immunological response to combination antiretroviral therapy (ART) among persons living with HIV-infection. The Johns Hopkins HIV Clinical Cohort, a prospective cohort, was used for this analysis. Data were collected every 6 months from before the initiation of antiretroviral treatment until the end of follow-up (2000-2008). Alcohol consumption was measured as the average quantity (number of drinks per drinking day) and frequency (number of days per week) of use in the prior 6 months. Immunological response was the dependent variable of interest, and was directly measured by CD4+ T cell count (a variable that can change over time). Alcohol consumption and time were assessed as a fixed effect; a random intercept and random slope were also specified for time. This allowed the investigators to assess intra-individual differences in baseline CD4 count (random intercept) and rate of change (random slope), as well as an average rate of change (fixed effect) in CD4 count over time. The investigators found that, overall there was no longitudinal association between alcohol consumption and CD4 count (fixed effect). See Sullivan *et al.* [32] and Crawford *et al.* [33] for additional examples of GLMM.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

3.3. Marginal Structural Models

Marginal structural models have the same structure as the marginal models upon which GEE is based, except that they are models for the causal effect of a time-dependent exposure for a counterfactual population in which the exposure variables are unaffected by previous confounding variables. Inverse-probability of exposure weights are used to match the sampled population, in which there is confounding, to this ideal counterfactual population, in which there is no confounding. Matching is based on the likelihood that a person would be in the "exposed" group or not, using information from other measured variables. Then a weighted GEE analysis is conducted. This methodology handles time-dependent confounding. Time-dependent confounders are simultaneously confounders of a subsequent exposure and intermediate variables on the pathway from a previous exposure to the outcome of interest. Using ordinary GEE or GLMM regression in which both alcohol treatment over time and alcohol consumption over time are included as covariates is subject to bias due to the inclusion of an intermediate variable in the regression model. Marginal structural models with inverse probability of exposure weighting overcome this problem.

As an example, Howe, Sander, Plankey, and Cole [3] sought to determine the association of alcohol consumption on HIV acquisition in a sample of injection drug users. Data from 1525 participants were available over a 10-year time period. The investigators realized that risky sex and drug use would be important confounding variables that could affect HIV acquisition, but that sexual behavior and drug use could be part of the causal pathway between alcohol and HIV acquisition. Therefore, using standard regression techniques to adjust for confounding could result in diminishing any association to a non-significant result. Marginal structural models used inverse weighting to deal with time-varying covariates. The investigators identified a statistically significant relationship of increased alcohol consumption with HIV acquisition, with evidence of a dose response. For comparisons, the investigators did the same analysis using more traditional methods to control for confounding variables. Although the results still indicated a trend of increased risk of HIV acquisition with more alcohol consumption, the magnitude of the association was lower (*e.g.* hazard ratio 1.7 instead of 2.2 for risk associated with >50 drinks/week), and the results were not statistically significant.

3.4. Time-Varying Effect Model

Another analytical approach for longitudinal data analysis is the time-varying effect model (TVEM) [34, 35], also known as a generalized additive model (GAM) [36] or a generalized additive mixed model (GAMM) [37], depending on whether random effects are used or not. These models are a special case of GEEs or GLMMs, in which the effects of time are modeled as smooth functions and time is interacted with the exposure variables. Even when modeling time-varying exposures on outcomes, some researchers tend to assume that the associations between the exposure and outcome are relatively stable over time. However, associations between an exposure and outcome may be different at one point in time compared to another. For example, the effect of alcohol consumption on risky sexual behavior could be stronger in adolescence compared to its effect in adulthood, even if the level of exposure to alcohol consumption is fairly stable over time. The focus of TVEM is on how the effect of an exposure on an outcome, at each time-point, changes overtime. The

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TVEM is suitable for multi-wave time-intensive data. In addition to time-varying predictors, time-invariant variables (*e.g.*, sex, race/ethnicity, country of origin) can also be analyzed. Similar to mixed effect modeling and GEE, subjects with missing data can also be included. The method has been used in tobacco research [38], but not in alcohol use research. A simulation study also reported the strength of the TVEM modeling in analyzing substance use data with ordinal response responses [34].

3.5. Developmental Trajectory Analyses

Developmental (group-based) trajectory analyses characterize patterns of exposure as well as outcomes, and can test the longitudinal relationship between exposure variables (either time-invariant or time-variant) with longitudinal outcome measures [39]. When variables change over time, one can consider the trajectory (pattern) of the variable over time, and the trajectory itself can be one way to describe a cumulative pattern of exposure within an individual or population. Other variables can influence the trajectory over time and in this case, the trajectory itself can be considered an outcome. Many types of data analysis options assume an average for an entire population; group-based trajectory modeling assumes that several distinct groups occur within a population that each share a common exposure or outcome pattern. Here we can identify distinctions between important subgroups of a population of interest that is more representative of the natural setting, as most populations are heterogeneous in nature. Group-based trajectory analysis allows one to model individual patterns that are then clustered into groups, for example, based on similarity of individual drinking patterns. A posterior probability that any one group-based trajectory adequately captures the individual patterns is also provided, with a probability of 0.7 being indicative of sufficient internal reliability [40]. Time-dependent covariates may influence the trajectory shape and variation around the average trajectory, whereas time stable variables influence group membership. Group-based trajectory analyses are very helpful when changes over time are not linear. Because the analysis allows one to describe linear, quadratic, cubic, etc. changes in a particular behavior over time, a trajectory can more accurately describe a change over time compared to a linear model.

Several limitations to the developmental trajectory analysis should be considered, including the somewhat subjective judgment for the specific number of trajectories and what type of changes (linear, quadratic, etc.) to specify. Ideally, one should assess the existing literature on the variable in question to understand the patterns that are already known to exist. An additional limitation of group-based trajectory analysis is that the final trajectories are based on an estimation of patterns, and cannot be interpreted as absolute patterns for all individual subjects within a trajectory grouping. Further, group-based trajectory modeling assumes that patterns within groups are homogenous, and that data are missing completely at random, and thus are more subject to bias with greater amounts of missing data. This method is also not suitable for variables that change rarely over time, such as stable chronic illnesses.

Marshall *et al.* [41] sought to identify long-term alcohol drinking trajectories among HIV-infected, sexually active MSM in a prospective cohort of veterans engaged in care in the United States. The investigators also sought to determine the socio-demographic, behavioral, and clinical correlates of membership in hazardous drinking trajectories in order to identify

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

modifiable factors that may improve alcohol treatment interventions for this population. The investigators used the AUDIT-C scores obtained over multiple time-points from 2002-2011 to conduct a group-based trajectory analysis. To determine the number of groups to model, the investigators relied on goodness of fit statistics, significance of trajectory shape coefficients, and posterior probabilities. Results of the group-based trajectory analysis revealed four distinct alcohol consumption risk groups: infrequent (accounting for 15.9%), low risk (36.5%), potentially hazardous (35.1%), and consistently hazardous (12.5%). The investigators used GEE to model the socio-demographic, behavioral, and clinical variables associated with being in the consistently hazardous drinking group. They concluded that financial insecurity and concurrent substance use were potentially modifiable predictors of consistently hazardous alcohol use. Alcohol drinking trajectories have also been examined among women with HIV [42].

3.6. Nonparametric Longitudinal Analyses

Although this review focuses on parametric and semiparametric longitudinal modeling approaches, sometimes the data clearly do not meet assumptions needed for parametric analysis (*e.g.* the data may have a skewed distribution at some timepoints but not others). Several approaches to nonparametric longitudinal data are discussed in the literature [43], and statistical software packages are adapting to provide nonparametric analysis options [44].

CONCLUSION

In summary, a growing number of longitudinal datasets are available for assessment of HIV-related outcomes, and to identify variables associated with these outcomes. In this paper, we discuss strategies to reduce bias when conducting analysis with existing data, and we provide examples from literature that includes a range of populations affected by HIV infection. Longitudinal analytic strategies should be used to reduce bias when exposures, confounding variables, or outcomes change over time. The choice of analytic method does not typically differ in relation to a specific study population of interest (*e.g.* MSM *vs* women with HIV). However, none of the analytic methods can address a selection bias in which the exposed population is selected in a different manner from the unexposed population. Several data analytic options are available to incorporate the changes in the amount, timing, or pattern of the exposure variable. Choosing the optimal analytic strategy should also consider whether the outcome of interest is binary or time-variant, and consider options to minimize analytic bias.

ACKNOWLEDGEMENTS

This work was supported by NIH grant U24AA022002 (P.I. RL Cook) and the University of Florida Graduate School Fellowship (PhD Student NE Kelso).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Biography

Robert L. Cook

REFERENCES

1. High KP, Brennan-Ing M, Clifford DB, et al. HIV and aging: state of knowledge and areas of critical need for research. A report to the NIH Office of AIDS Research by the HIV and Aging Working Group. J Acquir Immune Defic Syndr. 2012; 60(1):S1–18. [PubMed: 22688010]
2. Justice AC. HIV and aging: time for a new paradigm. Curr HIV/AIDS Rep. 2010; 7(2):69–76. [PubMed: 20425560]
3. Howe CJ, Sander PM, Plankey MW, Cole SR. Effects of time-varying exposures adjusting for time-varying confounders: the case of alcohol consumption and risk of incident human immunodeficiency virus infection. Int J Public Health. 2010; 55(3):227–8. [PubMed: 20143124]
4. Barkan SE, Melnick SL, Preston-Martin S, et al. The Women's Interagency HIV Study. WIHS Collaborative Study Group. Epidemiol Camb Mass. 1998; 9(2):117–25.
5. Justice AC, Dombrowski E, Conigliaro J, et al. Veterans Aging Cohort Study (VACS): Overview and description. Med Care. 2006; 44(8 Suppl 2):S13–24. [PubMed: 16849964]
6. Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, Rinaldo CR. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. Am J Epidemiol. 1987; 126(2):310–8. [PubMed: 3300281]
7. Singer, JD.; Willett, JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. 1st ed.. Oxford University Press; New York: 2003. p. 672
8. Conen A, Wang Q, Glass TR, et al. Association of alcohol consumption and HIV surrogate markers in participants of the swiss HIV cohort study. J Acquir Immune Defic Syndr. 2013; 64(5):472–8. [PubMed: 23892243]
9. Bryant KJ, Nelson S, Braithwaite RS, Roach D. Integrating HIV/AIDS and alcohol research. Alcohol Res Health J Natl Inst Alcohol Abuse Alcohol. 2010; 33(3):167–78.
10. Kelso NE, Sheps DS, Cook RL. The association between alcohol use and cardiovascular disease among people living with HIV: A systematic review. Am J Drug Alcohol Abuse. 2015; 30:1–10. [PubMed: 26225813]
11. Ronksley PE, Brien SE, Turner BJ, Mukamal KJ, Ghali WA. Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. BMJ. 2011; 342:d671. [PubMed: 21343207]
12. Fedirko V, Tramacere I, Bagnardi V, et al. Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. Ann Oncol Off J Eur Soc Med Oncol ESMO. 2011; 22(9):1958–72.
13. Neblett RC, Hutton HE, Lau B, McCaul ME, Moore RD, Chander G. Alcohol consumption among HIV-infected women: impact on time to antiretroviral therapy and survival. J Womens Health. 2011; 20(2):279–86.
14. De Stavola BL, Nitsch D, dos Santos Silva I, et al. Statistical issues in life course epidemiology. Am J Epidemiol. 2006; 163(1):84–96. [PubMed: 16306313]
15. Skinner HA, Sheu WJ. Reliability of alcohol use indices. The Lifetime Drinking History and the MAST. J Stud Alcohol. 1982; 43(11):1157–70. [PubMed: 7182675]
16. Bataller R, Brenner DA. Liver fibrosis. J Clin Invest. 2005; 115(2):209–18. [PubMed: 15690074]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

17. Fuster D, Tsui JI, Cheng DM, et al. Impact of lifetime alcohol use on liver fibrosis in a population of HIV-infected patients with and without hepatitis C coinfection. Alcohol Clin Exp Res. 2013; 37(9):1527–35. [PubMed: 23647488]
18. Chen WY, Rosner B, Hankinson SE, Colditz GA, Willett WC. Moderate alcohol consumption during adult life, drinking patterns, and breast cancer risk. JAMA. 2011; 306(17):1884–90. [PubMed: 22045766]
19. Bobashev GV, Liao D, Hampton J, Helzer JE. Individual patterns of alcohol use. Addict Behav. 2014; 39(5):934–40. [PubMed: 24569104]
20. Nelder J, Wedderburn R. Generalized Linear Models. J R Stat Soc. 1972; 135(3):370–84.
21. Liang K-Y, Zeger SL. Longitudinal Data Analysis Using Generalized Linear Models. Biometrika. 1986; 73(1):13–22.
22. Littell RC, Pendergast J, Natarajan R. Modelling covariance structure in the analysis of repeated measures data. Stat Med. 2000; 19(13):1793–819. [PubMed: 10861779]
23. Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. 2nd ed.. Wiley; New York, NY: 2002.
24. Pan W. Akaike's information criterion in generalized estimating equations. Biometrics. 2001; 57(1):120–5. [PubMed: 11252586]
25. Sullivan LE, Goulet JL, Justice AC, Fiellin DA. Alcohol consumption and depressive symptoms over time: a longitudinal study of patients with and without HIV infection. Drug Alcohol Depend. 2011; 117(2-3):158–63. [PubMed: 21345624]
26. Kahler CW, Wray TB, Pantalone DW, et al. Daily associations between alcohol use and unprotected anal sex among heavy drinking HIV-positive men who have sex with men. AIDS Behav. 2015; 19(3):422–30. [PubMed: 25194967]
27. Moriya A, Iwasaki Y, Ohguchi S, et al. Roles of alcohol consumption in fatty liver: A longitudinal study. J Hepatol. 2015; 62(4):921–7. [PubMed: 25433160]
28. Tsui JI, Cheng DM, Coleman SM, et al. Pain is associated with risky drinking over time among HIV-infected persons in St. Petersburg, Russia. Drug Alcohol Depend. 2014; 144:87–92. [PubMed: 25220898]
29. Seth P, Sales JM, DiClemente RJ, Wingood GM, Rose E, Patel SN. Longitudinal examination of alcohol use: a predictor of risky sexual behavior and Trichomonas vaginalis among African-American female adolescents. Sex Transm Dis. 2011; 38(2):96–101. [PubMed: 20739910]
30. Finucane MM, Samet JH, Horton NJ. Translational methods in biostatistics: linear mixed effect regression models of alcohol consumption and HIV disease progression over time. Epidemiol Perspect Innov EPI. 2007; 4:8.
31. Kowalski S, Colantuoni E, Lau B, et al. Alcohol consumption and CD4 T-cell count response among persons initiating antiretroviral therapy. J Acquir Immune Defic Syndr. 2012; 61(4):455–61. [PubMed: 22955054]
32. Sullivan LE, Saitz R, Cheng DM, Libman H, Nunes D, Samet JH. The impact of alcohol use on depressive symptoms in human immunodeficiency virus-infected patients. Addict Abingdon Engl. 2008; 103(9):1461–7.
33. Crawford TN, Sanderson WT, Breheny P, Fleming ST, Thornton A. Impact of non-HIV related comorbidities on retention in HIV medical care: does retention improve over time? AIDS Behav. 2014; 18(3):617–24. [PubMed: 23695522]
34. Dziak JJ, Li R, Zimmerman MA, Buu A. Time-varying effect models for ordinal responses with applications in substance abuse research. Stat Med. 2014; 33(29):5126–37. [PubMed: 25209555]
35. Tan X, Shiyko MP, Li R, Li Y, Dierker L. A time-varying effect model for intensive longitudinal data. Psychol Methods. 2012; 17(1):61–77. [PubMed: 22103434]
36. Hastie, TJ.; Tibshirani, RJ. Generalized Additive Models. Chapman & Hall; London: 1990.
37. Lin X, Zhang D. Inference in generalized additive mixed models by using smoothing splines. J R Statist Soc B. 1990; 61(2):381–400.
38. Shiyko MP, Lanza ST, Tan X, Li R, Shiffman S. Using the time-varying effect model (TVEM) to examine dynamic associations between negative affect and self confidence on smoking urges: differences between successful quitters and relapsers. Prev Sci Off J Soc Prev Res. 2012; 13(3): 288–99.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

39. Nagin DS. Group-based trajectory modeling: an overview. Ann Nutr Metab. 2014; 65(2-3):205–10. [PubMed: 25413659]

40. Andruff H, Carraro N, Thompson A, Gaudreau P. Latent class growth modelling: A tutorial. Tutor Quant Methods Psychol. 2009; 5(1):11–24.

41. Marshall BDL, Operario D, Bryant KJ, et al. Drinking trajectories among HIV-infected men who have sex with men: a cohort study of United States veterans. Drug Alcohol Depend. 2015; 148:69–76. [PubMed: 25596785]

42. Cook RL, Zhu F, Belnap BH, et al. Alcohol consumption trajectory patterns in adult women with HIV infection. AIDS Behav. 2013; 17(5):1705–12. [PubMed: 22836592]

43. Xiang D, Qiu P, Pu X. Nonparametric regression analysis of multivariate longitudinal data. Statistica Sinica. 2013; 23:769–789. Available at URL:/statistica/j23n2/j23n213/j23n213.html.

44. Noguchi K, Gel YR, Brunner E, Kimihiro F. nparLD: an R software package for the nonparametric analysis of longitudinal data in factorial experiments. J Stat Software. 2012; 50:12. Available at http://www.jstatsoft.org/v50/i12/paper.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Brief summary of advanced longitudinal analysis methods.

Method	Summary	Strengths	Limitations
Generalized Estimating Equation (GEE)	A marginal (population averaged) model and extends generalized linear modeling (GLM) by specifying correlations between repeated measures within subjects Can be used with time-varying outcome and time-invariant or time-varying exposure	Subjects do not need to have the same number of observations or have the same follow-up time-points and data can be missing at random for unbiased estimates	Produces fixed effects only, or population averaged parameters Not efficient to handle complex research designs
Mixed and Generalized Linear Mixed Model (GLMM)	Estimates the effect of an exposure within an individual as well as across a population Can be use with time-varying outcome and time-variant (fixed effect only) and time-varying (fixed and random effects) exposures	Subjects do not need to have the same number of observations or have the same follow-up time-points and data can be missing at random for unbiased estimates	Random effects are assumed to be independent of fixed effect covariates and normally distributed with constant variance
Marginal Structural Model	Models for the causal effect of a time-dependent exposure for a counterfactual population in which the exposure variables are unaffected by previous confounding variables Inverse-probability of exposure weights are used to match the sampled population, in which there is confounding, to this ideal counterfactual population, in which there is no confounding Can be used with time-varying outcome and time-varying exposures	Model complex causal relationship of factors related to alcohol use/treatment Deals with time-dependent confounding factors Most relevant for analyzing theory-based research	Confounders must cause the exposure variable Assumes that there are no unmeasured confounders
Time-Varying Effect Model	The effects of time are modeled as smooth functions and time is interacted with the exposure variables Measures change of the association between exposure and outcome over time Suitable for multi-wave, time-intensive data Can be use with time-varying outcome and time-variant (fixed effect only) and time-varying (fixed and random effects) exposures	Quantifies variations in the effect over time of a predictor on outcome Similar to GEE and GLMM, subjects with missing data during follow-ups can be included	Changes in effects over time are sensitive to changes in study sample due to missing data at each time point
Developmental/Group-based Trajectory	Characterizes distinct patterns of an exposure or outcome over time Assumes that several distinct groups occur within a population that each share a common exposure or outcome pattern Suitable for time-varying exposures and/or outcomes	Be able to handle data from a non-homogeneous population Allows specification of linear, quadratic, cubic, etc. changes in a particular behavior over time, a trajectory Can more accurately describe a change over time compared to a linear model	Partially relies on subjective judgment for number and type of trajectories Assumes that data are missing completely at random; more subject to bias with greater amounts of missing data