February, 2016

Transitioning from First Drug Use to Dependence Onset: Illustration of a Multiparametric Approach for Comparative Epidemiology

Olga A. Vsevolozhskaya, Michigan State University
James C. Anthony, Michigan State University

Available at: https://works.bepress.com/vsevolozhskaya/5/
Transitional from First Drug Use to Dependence Onset:

Illustration of a Multiparametric Approach for Comparative Epidemiology

Olga A. Vsevolozhskaya, James C. Anthony

Department of Epidemiology and Biostatistics, College of Human Medicine, Michigan State University, East Lansing MI 48824, USA

Corresponding Author:

James C. Anthony, Ph.D

Department of Epidemiology and Biostatistics

College of Human Medicine

Michigan State University

909 Fee Road

East Lansing, MI 48824-1030

janthony@msu.edu

Running Title: Multiparametric approach for comparative epidemiology

Word count: 3,730 (4,000 limit)
Abstract (236 words; 250 max)

Studying transitions from first drug use (DU) to drug dependence onset (DD), we estimate a parsimonious set of parameters based on epidemiological data, with plans for longitudinal research on pre-dependent newly incident drug users, tracking their self-administration frequencies and DD clinical features. Our expectation is a distinctive sigmoid pattern with one asymptote for lower DD probability (when DU is insubstantial), up-turning slopes of DD risk beyond a middle value (PD50), and eventual higher DD risk asymptotes at higher DU frequencies. We illustrate this novel approach using cross-sectional data from United States National Surveys on Drug Use and Health, 2002-2011. Empirical DD probabilities observed soon after newly incident use are estimated across drug use frequency values, using parametric Hill functions and four governing parameters for differential comparison across drugs and DU subgroups. Among the drug sub-types considered, cocaine showed larger estimates, especially among females (estimated Pmin = 7% for females vs. 3% for males; p<0.001), for whom PD50 is shorter by 8 days of use (p=0.031). Clear alcohol male-female differences were also observed (e.g., female PD50 < male PD50; p=0.002). Although we developed and refined this approach using cross-sectional snapshots soon after DU onset, this novel multiparametric statistical approach for comparative epidemiological drug dependence research creates new opportunities in planned studies with prospectively gathered longitudinal data. The cross-sectional estimates provide starting values needed to plan future longitudinal research programs intended to estimate effects of adolescent-onset drug self-administration.
Introduction

Within the field of epidemiological research on drug dependence in cross-sectionally observed community samples, there is a tradition of plotting the number of days of drug use on the x-axis and the estimated probability of being drug dependent on the y-axis (Apelberg et al, 2014; Chen et al, 1997; Chen and Kandel, 2002; Esser et al, 2014). The result is a figure that resembles a dose-response or dose-effect curve from laboratory experiments designed so that dosage level or number of dosage days are ‘fixed effect’ exogenous explanatory variables that are under the control of the experimenter. In truth, the variable on the x-axis in epidemiological research generally is an endogenous variable that is interdependent with the probability of being drug dependent, with likely feedback loops, as explained elsewhere (Anthony, 2010).

In this research project, we try to push this tradition of drug dependence epidemiology in a new direction, harnessing a functional analysis approach that can be extended from cross-sectional snapshots to dynamic longitudinal data. With focus on newly incident users (as opposed to prevailing or ‘prevalent’ users), our hope is to enhance understanding of heterogeneous published estimates, and with new estimates needed to guide longitudinal research on effects of adolescent-onset drug self-administration on later brain connectivity, as well as functional, structural, and sociobehavioral outcomes such as onset of drug dependence syndromes. Our primary research focus is the drug-by-drug comparison, but as explained below, we also illustrate how the approach can be used to study subgroups of drug users, with an example based on possible male-female differences.

At this stage in the research, we cannot solve the problem of feedback loops, but we can impose a partial constraint by focusing estimation on the first months after onset of newly incident drug use when feedback effects might be trivial. During these months, no more than
a small minority subset of drug users has developed patterns of sustained daily use that sometimes typify drug dependence cases.

To illustrate, in the first 1-2 years after onset of cocaine use, no more than about 5%-6% of cocaine users develop a syndrome of cocaine dependence. For cannabis and for alcohol, the corresponding transition probabilities are well below four percent (Wagner and Anthony, 2002). For this reason, in this project, we restrict the epidemiological samples to individuals for whom no more than 12 months has passed since first extra-medical use of the drug. Here, the adjective ‘extra-medical’ refers to drug use for feelings such as ‘to get high’ and otherwise using the drug outside boundaries that a prescribing clinician might intend in relation to approved indications and acceptable clinical practice.

By using the term ‘functional analysis’, we refer to a nonlinear statistical model based upon a known mathematical equation with parameters for relationships of interest. In the present study, we investigate occurrence of drug dependence syndromes as clinically meaningful responses as might be influenced by the frequency of drug-taking soon after initiation of extra-medical drug use. The relationship of interest is expected to be ‘S’-shaped (a sigmoid), as in most dose-response curves of pharmacology or toxicology. Our approach is to estimate a parsimonious set of epidemiologically meaningful parameters that characterize this ‘S’-shaped relationship, based on a nonlinear four-parameter Hill equation. Using this approach, we seek to simplify drug-by-drug comparisons in the comparative epidemiology of drug dependence, specifying our response variable to be the empirical probability of a drug dependence syndrome. The approach can be extended readily to other responses (e.g., different dependence features).

As distinct from concepts of $L_{D_{30}}$ and $E_{D_{30}}$, one of the Hill function parameters estimated under this model describes the estimated number of days of drug use corresponding to the
midpoint drug dependence location or half-probability of dependence, which we label as $PD_{10}$. The $PD_{10}$ estimate can be interpreted as a $\frac{1}{2}$-way point in the direction of an upper asymptote, $R_{\text{max}}$, where $R_{\text{max}}$ describes a limiting probability of becoming drug dependent when there is daily DU, as observed when the newly incident drug user is assessed (with an interval of up to 12 months since first use). The observed $R_{\text{max}}$ estimate for cocaine might approach 100%, but our expectation is that $R_{\text{max}}$ will have lower values for other drugs such as cannabis and alcohol. In some respects, the upper asymptote is akin to a ‘saturation point’ in pharmacokinetics data, possibly not attained within the range of the observed x-values (i.e., within the observed range of continuous daily DU). Nonetheless, these $PD_{10}$ estimates can still be derived using Hill equation fits and used for rank order analyses drug-by-drug (e.g., Engel et al, 2013; Frank, 2013; Prins et al, 1999).

Most epidemiological methods used to study sigmoid relationships have relied either on dose-response epidemiologic meta-analysis (Bagnardi et al, 2014; Berlin et al, 1993) or on one or more nonparametric techniques (Bagnardi et al, 2004; Royston, 2014; Royston and Sauerbrei, 2007). A restrictive meta-analysis assumption is that dose-response relationships are linear on the natural logarithm scale. The non-parametric approach suffers because its lower and upper asymptotes for the ‘S’-shaped curve are not limited to the (0,1) interval, whereas a practical model should allow neither negative probabilities of drug dependence for small amounts of drug use nor probabilities over 1 for large doses. Additionally, the nonparametric method essentially is a ‘connect the dots’ approach. It fails to offer a set of multiple parameter estimates we can use to compare one drug versus another.

Pharmaceutical scientists often use parametric functional analysis approaches (Ankomah and Levin, 2012; Black and Leff, 1983; Regoes et al, 2004; Zernig et al, 2007). The same is true for pre-clinical drug research (Ahmed and Koob, 2005) and for weed science (Brain and
Cousens, 1989; Seefeldt et al., 1995). Nonetheless, nonlinear functional modeling largely has been ignored in epidemiological field studies.

In order to draw attention to new lines of research opened up with this functional approach, we provide an illustration from our research on differences in the ‘S’-shaped curve for four different drug subtypes: alcoholic beverages (hereinafter, alcohol), cannabis, cocaine, and prescription opioid pain relievers (PPR). In the case of PPR, the domain of inquiry is restricted to extra-medical PPR use as defined above and elsewhere (Anthony et al., 1994). Readers might wish to know why we did not include tobacco products in this comparative analysis. The reason is that NSDUH assesses dependence on tobacco products only in relation to the 30 days prior to assessment. The result is an apples versus oranges comparison across drug subtypes. For tobacco, the focus is on dependence in past month, i.e., on the 30 days before assessment. For all other drugs, the focus is on dependence in past year, i.e., on the 12 months prior to assessment. There is no way to recalibrate a 30 day interval with a 12 month interval.

Additionally, pursuing a NIDA research agenda on male-female differences, we considered how the parameter estimates might differ for male and female drug users (Mello N.K., 1986; Sartor et al., 2014; Wagner and Anthony, 2007; Wetherington, 2007; Wilhelm et al., 2014). In prior work, female-male contrasts for risk of becoming dependent assume that age is held constant (e.g., Chen and Kandel, 2002) or hold constant elapsed time since first onset of drug use (e.g., Sartor et al., 2014; Wagner and Anthony, 2007). In contrast, here we explore male-female differences in chances of becoming drug dependent conditional on the same rate of drug use counted in relation to days of drug use in a specified interval prior to the date of assessment (i.e., the 30 day interval just prior to the date of survey assessment).
Materials and Methods

Population under Study, Sampling, and Measurements

The study population consists of non-institutionalized civilian residents of the United States (US), age 12 years and older, with a range of dwelling units that includes homeless shelters and other non-institutional group quarters, as well as households. Each year from 2002 through 2011, large independent multi-stage area probability samples of this study population were drawn for the National Surveys on Drug Use and Health (NSDUH; n > 65,000 each year). NSDUH protocols, approved by cognizant Institutional Review Boards, protected the human subject participants.

The main NSDUH measurement approach has involved completion of audio computer assisted self-interviews (ACASI), in either English or Spanish. Standardized multi-item modules have been used to assess broad range of drug use and health indicators.

The module items on drug use and drug dependence, with criteria from the penultimate Diagnostic and Statistical Manual (DSM-IV; American Psychiatric Association, 2013), are available online http://www.epi.msu.edu/vsevoloz/scripts/Hill_function/html_files/. These and other details about the NSDUH have been published in multiple prior journal articles (e.g., Seedall and Anthony, 2013; Vsevolozhskaya and Anthony, 2014) and are described in readily available online publications (http://oas.samhsa.gov/nsduh/methods.cfm).

Data Source

Due to confidentiality concerns, non-governmental researchers typically do not have direct access to the NSDUH restricted-use micro-data (http://www.icpsr.umich.edu/icpsrweb/content/SAMHDA/dataportal.html). For the present investigation, we fit multiparametric models to data from tabulated summary statistics from
the NSDUH Restricted-use Data Analysis System (R-DAS), based on fieldwork completed between 2002 and 2011 (http://www.icpsr.umich.edu/icpsrweb/content/SAMHDA/rdas.html). (In 2016, once NSDUH data from 2012 through 2015 become available, cross-replication of this study’s estimates will become possible.)

Statistical analysis: Focus on newly incident drug users

Within these cross-sectional R-DAS datasets there are month-by-month histories about drug use initiation, making it possible to identify ‘newly incident drug users’ or ‘recent initiates’ (i.e., those for whom no more than 12 months have passed between initiation of use and the date of survey assessment). The assessment indicates which newly incident users developed DSM-IV dependence within that first year after initiation, as well as frequencies of drug use. The list of the R-DAS variables used to generate tabulated estimates and standard errors is available online at http://www.epi.msu.edu/vsevoloz/scripts/Hill_function/html_files/ . R-DAS tabulated data presented in a user-friendly format is available at http://www.epi.msu.edu/vsevoloz/scripts/Hill_function/txt_files/ . Standard errors are from Taylor series expansion, as described by Heeringa et al, (2010).

Statistical analysis: Hill model-fitting

We fit ‘S’-shaped curves to model relationships linking frequency of drug use in the month just prior to assessment and the probability of developing DD within the first year after DU initiation. The mathematical equation describing this sigmoid relationship is a Hill function:

\[ y = (P_{\text{max}} - P_{\text{min}}) \times \frac{1}{\left( 1 + \left( \frac{P_{\text{max}}}{P_{\text{min}}} \right)^n \right)} + P_{\text{min}} \quad (1) \]
where \( y \in [0, 1] \) is the DD incidence rate, \( x = 0, 1, \ldots, 30 \) is the frequency of drug use as governed by the number of days of drug use in the month just prior to the NSDUH interview date, \( P_{\text{min}} \) is the lower asymptote, \( P_{\text{max}} \) is the upper asymptote, \( k \) is the Hill coefficient that measures the steepness of the curve, and \( PD_{50} \) denotes half probability of dependence, \( (P_{\text{max}} - P_{\text{min}})/2 \). Figure 1 provides geometric interpretation of the parameters.

The mathematical expression in Equation (1) approximates an ‘ideal’ situation from a deterministic model in which both the predictor values \( x \) and the response values \( y \) are observed without errors and are ‘true’. In practice, however, the observed survey data come with measurement errors and only approximations of true population values. Combining all measurement errors into a single term \( \epsilon \), we can rewrite Equation (1) as

\[
y = f(x; (P_{\text{min}}, P_{\text{max}}, k, PD_{50}))+\epsilon
\]

Equation (2) is an example of a nonlinear statistical regression model, because the response is a nonlinear function of the unknown parameters \( (P_{\text{min}}, P_{\text{max}}, k, PD_{50}) \). This nonlinear model can be fit using a standard least squares minimization algorithm implemented in multiple statistical software packages, e.g., nls() R function (http://www.r-project.org/). For readers not familiar with R statistical software, we refer to Ritz and Streibig text (2008). We also are sharing our work with others via online access to our R scripts for parameters \( (P_{\text{min}}, P_{\text{max}}, k, PD_{50}) \) (http://www.epi.msu.edu/vsevoloz/scripts/Hill_function/R_scripts/).

Finally, because this work is a ‘proof of concept’ illustration at this stage, we did not model multiple covariates or suspected influences such as earlier age-of-onset variations we intend to incorporate in future research. The contrasts of male and female subgroups serve to illustrate an approach that can be used when estimating these effects.
Results

Development of Drug Dependence Syndrome

Nonlinear regression approaches require users to start by supplying a set of possible starting values of the parameters (i.e., $P_{\min}$, $P_{\max}$, $k$, $PD_{50}$). For starting values in this project, we produced ‘eye-ball’ estimates while plotting R-DAS-generated DD rate estimates on the y-axis and the count of days of DU on the x-axis. For all four drug subtypes, we then derived estimates of the risk of developing DD syndrome by recent frequency of DU with corresponding 95% confidence intervals (CIs), as shown in Figure 2. As it happens, few newly incident users progressed to frequent DU within the first year, such that observed 95% confidence intervals (CI) became quite wide across increasing values of the count of DU days. We took this condition into account using inversed empirical variances as weights. As noted above, we also restricted parameter values of $P_{\min}$ and $P_{\max}$ to be in the allowable closed interval (0,1). As shown in Figure 2, for all drug subtypes, the estimated probability of DD among newly incident drug users was observed to be larger across levels of increasing recent DU frequency, with eventual leveling off.

Comparison of the Estimated Parameters Across Drugs

Table 1 shifts focus from qualitative comparison of ‘S’-shaped curves toward the four estimated parameters from the Hill equation, $P_{\min}$, $P_{\max}$, $PD_{50}$, and $k$, with 95% weighted residual bootstrap confidence intervals (CIs). Supplementary information accompanying this paper provides a detailed discussion of the weighted bootstrap procedure we used to obtain the CIs.
In terms of the four estimated parameters, one might expect cocaine to have the largest estimated values of $P_{\min}$, $P_{\max}$, $PD_{20}$, and $k$ among the drug subtypes considered, given cocaine’s exceptional functional value as a reinforcer and previously observed epidemiological estimates (e.g., Wagner and Anthony, 2002). The observed estimates are consistent with this expectation. For those who used cocaine at least once in the past year but with no use in the month of the interview, i.e. $x = 0$, the estimated probability of having become dependent is 5% (95% confidence interval, CI = 4%, 6%). Nonetheless, the $PD_{20}$ mid-value is estimated as 18 days of recent cocaine use (95% CI = 9, 27 days). For the very small subset of daily users, the DD probability is essentially one (100%).

By comparison, parameter estimates for newly incident alcohol users are dampened, but noteworthy. Among newly incident drinkers with no recent drinking, the estimated probability of alcohol dependence (AD) is 1%; alcohol’s $PD_{20}$ estimate is 4 days of recent drinking (95% CI = 3, 10 days). Among those consuming alcohol essentially every day, an estimated 10% qualified as AD cases (95% CI = 6%, 17%).

We also present estimates for prescription pain relievers (generally opioid PPR) and for cannabis, which resemble one another. We note however that the estimates for PPR are based on relatively small numbers of days of drug use and small number of cases. The result is low statistical precision and very wide confidence intervals.

**Female-Male Contrasts**

Figure 3 depicts estimated female-male differences for alcohol dependence probabilities as a function of recent drinking days. The contrasting Hill equation parameter estimates are not too distant from one another with one exception, $PD_{20}$. Conditional on the same rate of
recent alcohol use, the estimated $P_{D_{20}}$ is 7 days for females and 11 days for males (p-value for the difference = 0.002).

Figure 4 shows estimates for cocaine dependence. Relative to males, newly incident cocaine-using females are more likely to develop cocaine dependence soon after first cocaine use, even among the lowest frequency newly incident users ($P_{\text{min}} = 7\%$ for females vs. 3\% for males; p-value <0.001). With frequency of cocaine use held constant, women seem to be more likely to become newly incident cases of cocaine dependence. The estimated $P_{D_{20}}$ is 14 days of recent use for females, and 22 days for males (p-value for the difference = 0.027).

For cannabis and prescription pain relievers, no statistically significant differences can be seen (p > 0.05), even though the numbers of newly incident users are substantial. For the PPR, one explanation for lack of statistical significance at the conventional level in this female-male contrast is the relatively smaller numbers of newly incident PPR users who are becoming dependent with the first year after onset of extra-medical PPR use.

**Discussion**

By turning to functional analysis of Hill equations, we offer a novel approach not previously seen in drug dependence epidemiology’s drug-by-drug comparisons. Based on theory and prior studies, our approach assumes that clinically significant dependence syndromes generally do not appear suddenly. Many extra-medical drug users try the drug once or a few times, and then never use the drug again. In other users, the dependence syndrome takes form in the days and months after initial drug use, with incremental strengthening and eventually leveling off of the probability of becoming drug dependent, with a resulting ‘S’-shaped curve resembling dose-response curves of pharmacology and toxicology. The Hill function yields four
parameter estimates that help enable drug-by-drug comparisons not previously seen in epidemiological research.

How well do our Hill equation parameter estimates for newly incident drug users replicate what others have found in prior studies with samples of all prevailing users (i.e., when no distinctions are drawn between those who just started to self-administer versus those who have been using the drug for more than one year). For cocaine, our $R_{min}$ estimate for the probability of cocaine dependence being seen among newly incident users with low DU frequency is 5%, not too distant from the 5%-6% estimate observed elsewhere in the first 1-2 years after first cocaine use, but substantially lower than the estimated 16%-20% value observed when cocaine users are studied many years after first cocaine use, irrespective of DU frequency (Reboussin and Anthony, 2006; Wagner and Anthony, 2002). Studying more recent epidemiological samples of cocaine users, Lopez-Quintero et al, (2011) produced a corresponding estimate of 20%-21% for ever-users of cocaine, but that study’s approach failed to include lower frequency cocaine users. Studying a prevalence sample of daily cocaine users, Chen and Kandel (2002) found that 74% were cocaine dependent.

With respect to alcohol, the life-table approach used by Wagner-Anthony indicated that as many as 12%-13% of drinkers develop alcohol dependence within the first 10 years after onset of drinking (Wagner and Anthony, 2002). The Lopez-Quintero study (Lopez-Quintero et al, 2011) estimated that 22%-23% of ever-drinkers had developed AD, but here again, the approach excluded low frequency drinkers. In Ridenour et al, (2006), the estimate of active AD among daily drinkers is 15%. This study’s Hill function estimate for newly incident daily drinkers is 10%, not too distant from the Ridenour estimate for all daily drinkers identified in a prevalence sample (i.e., with no restriction to newly incident drinkers).
Turning to cannabis, we note that Wagner and Anthony (2002) found that roughly 8% had become cannabis dependent within 10 years after first cannabis smoking. The Lopez-Quintero study estimate was 9%, even though low frequency cannabis smokers were excluded (Lopez-Quintero et al., 2011). Estimates from Chen et al (2005) suggest that four percent of newly incident cannabis users develop cannabis dependence within 24 months after initiation, with somewhat larger values seen among adolescent-onset users. Studying a prevalence sample of adolescents using cannabis daily, Chen et al (1997) found that 30% were cannabis dependent. More tightly focused on cannabis users studied within 12 months of first use, this study produced a corresponding Hill function estimate for $P_{\text{min}}$ of 2%, and for daily cannabis users estimated $P_{\text{max}}$ is 36%, irrespective of age.

This study’s $P_{\text{min}}$ and $P_{\text{max}}$ values for newly incident extra-medical users of prescription pain relievers are, respectively, 4% and 38%. The corresponding PPR estimate from the Wagner-Anthony life-table approach was 9% for all users, based on experiences accumulated during the first 10-15 years after initiation (Wagner and Anthony, 2002). Studying all prevalent PPR users cross-sectionally, Becker et al, (2008) and Novak et al, (2009) produced estimates of 13% and 20% for the probability of opioid dependence (OD) among current PPR users. Martins et al, (2007) estimated the odds of opioid dependence for frequent PPR users (50+ days in the past year) and the odds of OD for infrequent PPR users (1 day in the past year), and derived an odds ratio estimate of 14. The corresponding OR based on our Hill function analysis estimates of $P_{\text{min}}$ and $P_{\text{max}}$ for newly incident PPR users is 15.

As for female-male differences, there is a relatively limited epidemiology trace of published estimates (e.g, Bobzean et al, 2014; Chen and Kandel, 2002; Kasperski et al, 2011; Sartor et al, 2014). Overall, for cocaine, these investigations suggest that women, in the past, have been less likely to become newly incident cocaine users, and that there has been no
pronounced female-male difference in the probability of transitioning from initial cocaine use into the clinically significant state of cocaine dependence (e.g., see Wagner and Anthony, 2007). For alcohol, there is some evidence of a more rapid transition from alcohol onset until alcohol dependence for females (e.g., see Hernandez-Avila et al, 2004; Mello N.K., 1986; Zilberman et al, 2003). For cannabis, the female-male contrasts for these estimates are mixed (e.g., Chen et al, 1997; Hernandez-Avila et al, 2004). More research on female-male differences is needed for all of these drug compounds, especially for cannabis and for PPR, where the available published evidence is spotty.

Readers might wish to know why we did not include tobacco products in this comparative analysis. The reason is that NSDUH assesses dependence on tobacco products only in relation to the 30 days prior to assessment. The result is an ‘apples versus oranges’ comparison across drug subtypes. For tobacco, the focus is on dependence in past month, i.e., on the 30 days before assessment. For all other drugs, the focus is on dependence in past year, i.e., on the 12 months prior to assessment. There is no way to recalibrate a 30-day interval with a 12-month interval.

Before any further discussion of the Hill model for comparative epidemiological research on this topic, several study limitations deserve attention. Of special concern is the self-report interview data from NSDUH. Most likely, in the context of nationally representative sample surveys on this scale, there are few logistically feasible and affordable alternatives to self-report. Counter-balancing our acknowledgment of limitations in self-report interview data, there is a body of evidence indicating that self-report assessment methods can have reasonable levels of reliability and validity (Del Boca and Darkes, 2003; Vignali et al, 2012). Additionally, the NSDUH team conducted detailed methods studies on these topics (http://www.samhsa.gov/data/sites/default/files/2k6ReliabilityP/2k6ReliabilityP.pdf). These methodological studies do not indicate perfection of these assessments, but the
findings from the methods research have been generally supportive of reliability and validity of survey responses.

In addition, cross-sectional data always will be inferior to longitudinal data when fitting Hill functions of this type. Nonetheless, as noted in our introduction, this study’s cross-sectional estimates can serve as starting values for future longitudinal analyses, and should help guide newly planned longitudinal research on this topic. To illustrate, from this cross-sectional work, it can be anticipated that exceptionally large samples of newly incident PPR users will be required for effective estimates of Hill function parameters in the longitudinal context. Furthermore, within 1-2 years, it will be possible to seek replication of the cross-sectional estimates, once newer NSDUH data from surveys in 2012-2015 become available. We note that replication of large sample longitudinal study estimates is rare, and makes one wonder whether large-sample longitudinal study estimates ever will be replicated with independent samples. Nonetheless, cross-sectional study estimates might be of value while we await systematic replication of longitudinal study estimates because it always will be possible to compare estimates from cross-sectional studies with those obtained via longitudinal research designs.

One additional limitation of cross-sectional data might be avoided by asking newly incident drug users to give a month-by-month, or perhaps a week-by-week report about frequency of use and occurrence of clinical features, week by week since onset of first use. In this way, it might be possible to tease apart paths that run from frequency of use toward onset of clinical features, versus paths running from clinical features toward frequency of use. To the best of knowledge, no research project has attempted this intensity of measurement about fine-grained time-sequenced drug experiences. It remains on the agenda for potential future research.
Conceptualizing extensions beyond this initial study, we propose that one way to accelerate this line of research is to formalize a multivariate vector of drug dependence clinical features, each of which represents a constituent unit of response to drug exposure. We then can substitute this multivariate response vector in place of standardized diagnosis of drug dependence of the type considered in this illustration. An advantage of this substitution is that there is a larger number of newly incident users who experience subjectively felt tolerance or time displacement of other valued non-drug activities as the drug user starts to ‘spend more time’ on drug activities, as compared to the number who qualify as diagnosable cases of drug dependence. Moreover, when the constituent clinical features of DD are studied as responses, an allowance can be made for the possibility of a shift of the response curve to the left for clinical features that take form quite quickly after onset of extra-medical use (e.g., subjectively felt tolerance; ‘displacement’ or ‘spending more time [with the drug]’) and a corresponding shift to the right for clinical features that take more time to appear (e.g., manifestations of withdrawal upon abrupt discontinuation of use). Ultimately, elaboration of the modeling process to allow for heterogeneity of parameter estimates across clinical features will be useful. Here again, estimates based on already available cross-sectional data can be used to frame specific hypotheses against non-null alternatives that otherwise would be constrained to tests for departures from the null.

Other forms of heterogeneity in parameter estimates should be anticipated beyond this work’s initial focus on drug-by-drug variations and male-female differences. To illustrate another source of heterogeneity, recent studies (e.g., Kandel and Kandel, 2014) suggest that an allowance should be made for heterogeneity of cocaine dependence processes in relation to ‘pre-treatment’ with nicotine products such as tobacco cigarettes or smokeless tobacco, with the minority of newly incident cocaine users having no past nicotine exposures showing different parameter estimates than the majority of newly incident cocaine users who
generally have started to smoke tobacco and use nicotine products prior to onset of cocaine use. Even though our project has not taken into account heterogeneity introduced when multiple drugs are used, this initial work with readily available cross-sectional data should have future impact on the field of ‘polydrug use research’ (Anthony et al., in press), as manifest in recent NIDA program announcements to promote research on male-female differences of the type seen here when each drug compound has been considered one at a time, without reference to ‘pre-treatment’ effects or concurrent and simultaneous drug use.
Funding and Disclosure

This work is supported by a NIDA T32 research training program grant award (T32DA021129) for OV’s postdoctoral fellowship, by JCA's NIDA Senior Scientist and Mentorship Award (K05DA015799), and by Michigan State University. Otherwise, authors have nothing to disclose.
References


Table 1: Parameter Estimates and with the 95% Bootstrap Confidence Intervals. Data from United States National Surveys on Drug Use and Health, 2002-2011.

<table>
<thead>
<tr>
<th>Parameters (95% Bootstrap Confidence Intervals)</th>
<th>Parameters (95% Bootstrap Confidence Intervals)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannabis</td>
<td>P_{min} 0.02 (0.01, 0.02) P_{max} 0.36 (0.27, 0.71) P_{D_{90}} 11 (6, 33)* k 1.50 (1.07, 2.01)</td>
</tr>
<tr>
<td>PPR</td>
<td>P_{min} 0.04 (0.00, 0.05) P_{max} 0.38 (0.19, 1.00) P_{D_{90}} 11 (5, 80)* k 1.81 (0.80, 5.93)</td>
</tr>
<tr>
<td>Alcohol</td>
<td>P_{min} 0.01 (0.01, 0.01) P_{max} 0.10 (0.06, 0.17) P_{D_{90}} 4 (3, 10) k 2.48 (1.47, 6.17)</td>
</tr>
<tr>
<td>Cocaine</td>
<td>P_{min} 0.05 (0.04, 0.06) P_{max} 0.97 (0.46, 1.00) P_{D_{90}} 18 (9, 27) k 1.82 (1.31, 3.57)</td>
</tr>
</tbody>
</table>

*Since the bootstrap procedure yields non-symmetrical CIs, the upper bound of the estimate falls outside of the observed range of at most 30 days of continuous drug use prior to the assessment time. This phenomenon reflects that upper bound is not always attained within an interval of 30 days.
Figure legends

Figure 1: The ‘S’-shaped curve with the count of days of drug use on the x-axis and estimated risk of developing drug dependence on the y-axis. The four parameters that determine the shape of the curve are $P_{\min}$, $P_{\max}$, $P_{FU}$, and $h$. Data from newly incident drug users found in the United States National Surveys on Drug Use and Health, 2002-2011.

Figure 2: Dots and vertical bars are observed empirical estimates for risk of drug dependence with the corresponding 95% confidence intervals. Lines show estimated probability of having developed drug dependence across drug frequency strata, as estimated via nonlinear regression, and shaded regions show corresponding 95% bootstrap confidence intervals. Data from newly incident drug users found in the United States National Surveys on Drug Use and Health, 2002-2011.

Figure 3: Female-Male contrasts in estimated probability of alcohol dependence across levels of recent drinking. Data from newly incident drug users found in the United States National Surveys on Drug Use and Health, 2002-2011.

Figure 4: Female-Male contrasts in estimated probability of cocaine dependence across levels of recent cocaine use. Data from newly incident drug users found in the United States National Surveys on Drug Use and Health, 2002-2011.
Risk of developing drug dependence

$P_{max}$

$P_{min}$

Slope ($k$)

PD50

Number of days or occasions of drug use (on the log scale)