Neuropsychopharmacology (2016) 41, 869–876

www.neuropsychopharmacology.org

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

Olga A Vsevolozhskaya[1] and James C Anthony*,[1]

[1]*Department of Epidemiology and Biostatistics, College of Human Medicine, Michigan State University, East Lansing, MI, USA*

Studying transitions from first drug use (DU) to drug dependence (DD) onset, we estimate a parsimonious set of parameters based on epidemiological data, with plans for future longitudinal research on newly incident drug users and with tracking of 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), upturning slopes of DD risk beyond a middle value (PD_{50}), and eventual higher DD risk asymptotes at higher DU frequencies. We illustrate this novel approach using cross-sectional data from the United States National Surveys on Drug Use and Health, 2002–2011. Empirical DD probabilities observed soon after newly incident use are estimated across DU frequency values, using parametric Hill functions and four governing parameters for differential comparison across drugs and DU subgroups. Among drug subtypes considered, cocaine shows larger estimates, especially among females (estimated $P_{min} = 7\%$ for females *vs* 3% for males; $p < 0.001$), for whom PD_{50} is shorter by 8 days of use ($p = 0.027$), conditional on the same rate of use in the past 30 days. Clear alcohol male–female differences also are observed (eg, female PD_{50} < male PD_{50}; $p = 0.002$). Although based on cross-sectional snapshots soon after DU onset, this novel multiparametric statistical approach for comparative epidemiological DD 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 on transitions from initial DU until formation of a DD syndrome.
Neuropsychopharmacology (2016) **41,** 869–876; doi:10.1038/npp.2015.213; published online 12 August 2015

INTRODUCTION

Within the field of epidemiological research on drug dependence (DD) in cross-sectionally observed community samples, there is a tradition of plotting the number of days of drug use (DU) 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 DD 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 to present new estimates needed to influence study designs and to guide future studies. Our primary research focus is the drug-by-drug comparison, but as explained below, we also illustrate how the approach might 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 DU 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 DD cases.

To illustrate, in the first 1–2 years after onset of cocaine use, no more than ~5%–6% of cocaine users develop a syndrome of cocaine dependence. For cannabis and for alcohol, the corresponding transition probabilities are well below 4% (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 DU for feelings such as 'to get high' and otherwise using the drug outside boundaries that a

*Correspondence: Dr JC Anthony, Department of Epidemiology and Biostatistics, College of Human Medicine, Michigan State University, 909 Fee Road, East Lansing, MI 48824-1030, USA, Tel: +1 517 353 8623 100, Fax: +1 517 432 1130, E-mail: janthony@msu.edu
Received 27 January 2015; revised 30 June 2015; accepted 1 July 2015; accepted article preview online 15 July 2015

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 on a known mathematical equation with parameters for relationships of interest. In the present study, we investigate occurrence of DD syndromes as clinically meaningful responses as might be influenced by the frequency of drug taking soon after initiation of extra-medical DU. 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 DD, specifying our response variable to be the empirical probability of a DD syndrome. The approach can be extended readily to other responses (eg, different dependence features).

As distinct from concepts of LD_{50} and ED_{50}, one of the Hill function parameters estimated under this model describes the estimated number of days of DU corresponding to the midpoint DD location or half-probability of dependence, which we label as PD_{50}. The PD_{50} estimate can be interpreted as a half-way point in the direction of an upper asymptote, P_{max}, where P_{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 P_{max} estimate for cocaine might well approach 100%, but our expectation is that P_{max} might 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 (ie, within the observed range of continuous daily DU). Nonetheless, these PD_{50} estimates can still be derived using Hill equation fits and used for rank-order analyses drug-by-drug (eg, 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 nonparametric 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 DD for small amounts of DU nor probabilities over 1 for large doses. In addition, 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 preclinical 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 (PPRs). In the case of PPRs, the domain of inquiry is restricted to extra-medical PPR use as defined above and elsewhere (Anthony *et al*, 1994).

In addition, pursuing general NIH program interest in male–female differences, we considered how parameter estimates might differ for male and female drug users (Mello, 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 (eg, see Chen and Kandel, 2002) or hold constant elapsed time since first onset of DU (eg, see 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 DU counted in relation to days of DU in a specified interval before the date of assessment (ie, the 30-day interval just before 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, 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, in either English or Spanish. Standardized multi-item modules have been used to assess broad range of DU and health indicators.

The module items on DU and DD, 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 (eg, Seedall and Anthony, 2013; Vsevolozhskaya and Anthony, 2014; Ryan *et al*, 2012; Vaish *et al*, 2013) and are described in readily available online publications (http://oas.samhsa.gov/nsduh/methods.cfm).

Data Source

Owing 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 have become available, cross-replication of this study's estimates will become possible. (For this reason, to create an opportunity for independent replication, now we do not use available data from 2012 to 13 in this research report.)

Statistical Analysis: Focus on Newly Incident Drug Users

Within these cross-sectional R-DAS data sets, there are month-by-month histories about DU initiation, making it possible to identify 'newly incident drug users' or 'recent initiates' (ie, 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 DU. The list of the R-DAS variables used to generate tabulated estimates and SEs 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 are available at http://www.epi.msu.edu/vsevoloz/scripts/Hill_function/txt_files/. SEs 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 DU in the month just before 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:

$$y = (P_{\max} - P_{\min}) \cdot \frac{1}{1 + \left(\frac{\mathrm{PD}_{50}}{x}\right)^{k}} + P_{\min}, \qquad (1)$$

where $y \in [0, 1]$ is the DD incidence rate, $x = 0, 1, \ldots, 30$ is the frequency of DU as governed by the number of days of DU in the month just before the NSDUH interview date, $P_{\min}$ is the lower asymptote, $P_{\max}$ is the upper asymptote, k is the Hill coefficient that measures the steepness of the curve, and PD_{50} denotes half probability of dependence, $(P_{\max}-P_{\min})/2$. 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 ϵ, we can rewrite equation (1) as

$$y = f(x, (P_{\min}, P_{\max}, k, \mathrm{PD}_{50})) + \epsilon. \qquad (2)$$

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

Finally, as this work is a 'proof-of-concept' illustration at this stage, we have not modeled multiple covariates or suspected influences such as age-of-onset variations as might be incorporated in future research. The contrasts of male and female subgroups serve to illustrate an approach that can be used when estimating these suspected effects.

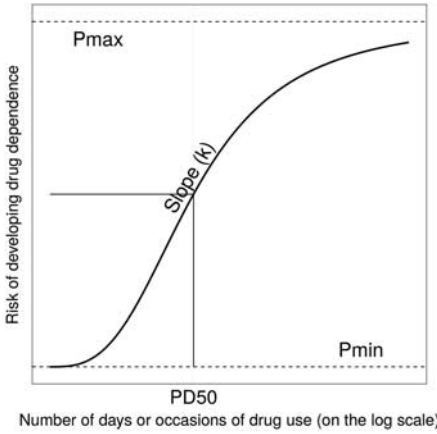

Figure 1 The 'S'-shaped curve with the count of days of drug use on the x axis and estimated risk of developing drug dependence on the y axis. The four parameters that determine the shape of the curve are $P_{\min}$, $P_{\max}$, PD_{50} and k. Data from newly incident drug users found in the United States National Surveys on Drug Use and Health, 2002–2011.

RESULTS

Development of DD Syndrome

Nonlinear regression approaches require users to start by supplying a set of possible starting values of the parameters (ie, $P_{\min}$, $P_{\max}$, k, and PD_{50}). For starting values in this project, we produced 'eyeball' 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% CIs 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

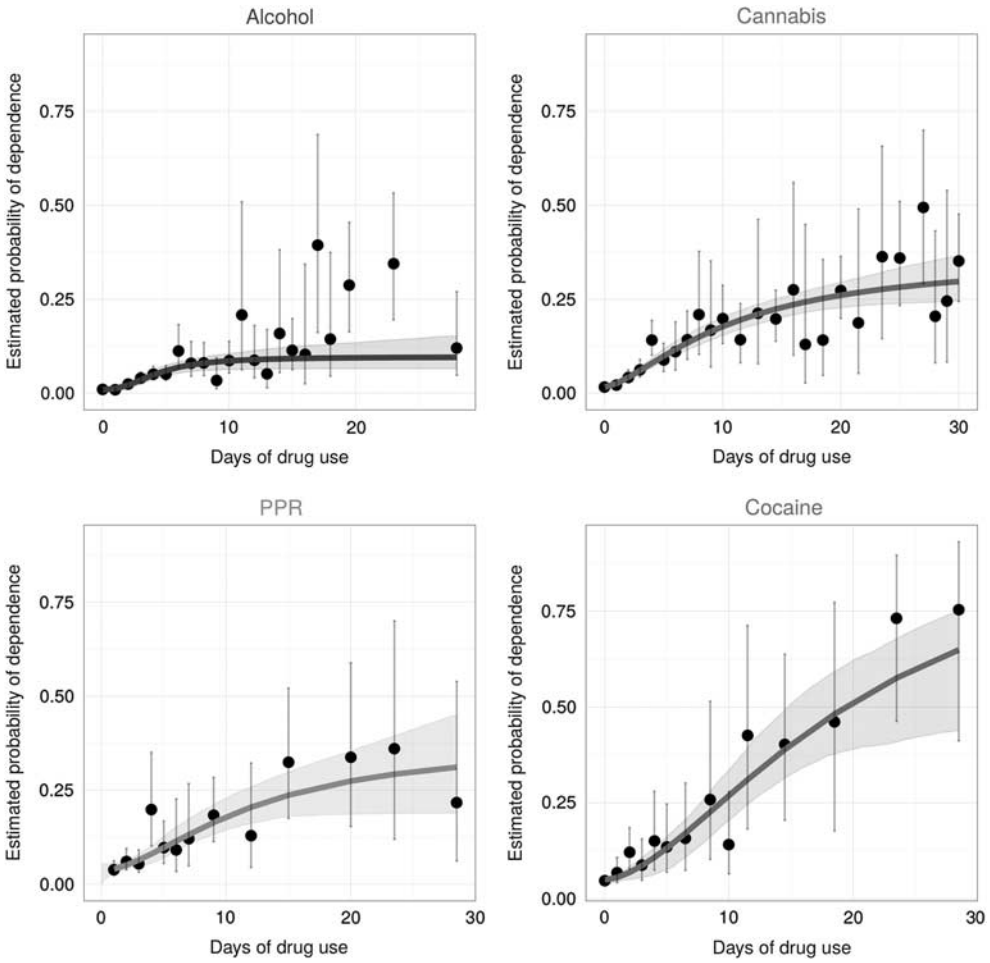

Figure 2 Dots and vertical bars are observed empirical estimates for risk of drug dependence (DD) with the corresponding 95% confidence intervals (95% CI). Lines show estimated probability of having developed DD across drug-frequency strata, as estimated via nonlinear regression, and shaded regions show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

Table 1 Parameter Estimates and with the 95% Bootstrap CIs

	Parameters (95% bootstrap CIs)			
	P_{min}	P_{max}	PD_{50}	k
Cannabis	0.02 (0.01, 0.02)	0.36 (0.27, 0.71)	11 (6, 33)[a]	1.50 (1.07, 2.01)
PPR	0.04 (0.00, 0.05)	0.38 (0.19, 1.00)	11 (5, 80)[a]	1.81 (0.80, 5.93)
Alcohol	0.01 (0.01, 0.01)	0.10 (0.06, 0.17)	4 (3, 10)	2.48 (1.47, 6.17)
Cocaine	0.05 (0.04, 0.06)	0.97 (0.46, 1.00)	18 (9, 27)	1.82 (1.31, 3.57)

Abbreviations: CI, confidence interval; PPR, prescription opioid pain reliever.
Data from the United States National Surveys on Drug Use and Health, 2002–2011.
[a]As the bootstrap procedure yields non-symmetrical CIs, the upper bound of the PD_{50} estimate falls outside of the observed range of at most 30 days of continuous drug use before the assessment time. This phenomenon reflects that P_{max} upper bound is not always attained within an interval of 30 days.

residual bootstrap CIs. Supplementary Information accompanying this paper provides a detailed discussion of the weighted bootstrap procedures 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_{50} and k among the drug subtypes considered, given cocaine's exceptional functional value as a reinforcer and previously observed epidemiological estimates (eg, see 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, ie $x=0$, the estimated probability of having become dependent is 5% (95%CI = 4%, 6%). Nonetheless, the PD_{50} 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 1 (100%).

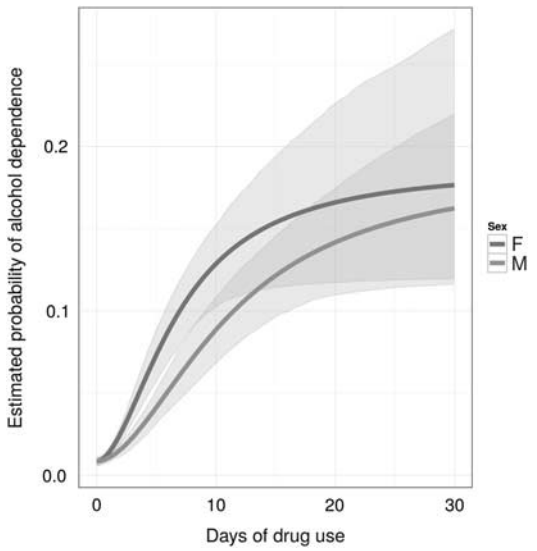

Figure 3 Female–male contrasts in estimated probability of alcohol dependence across levels of recent drinking. Data from newly incident drug users found in the United States National Surveys on Drug Use and Health, 2002–2011.

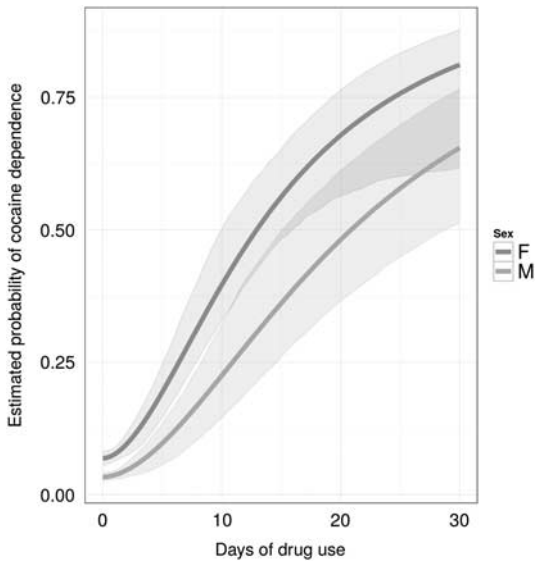

Figure 4 Female–male contrasts in estimated probability of cocaine dependence across levels of recent cocaine use. Data from newly incident drug users found in the United States National Surveys on Drug Use and Health, 2002–2011.

By comparison, parameter estimates for newly incident alcohol users are smaller but noteworthy. Among newly incident drinkers with no recent drinking, the estimated probability of alcohol dependence (AD) is 1%; alcohol's PD_{50} estimate is 4 days of recent drinking (95% CI = 3, 10 days). Among newly incident users 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 PPRs) and for cannabis, which resemble one another. We note however that the estimates for PPR are based on relatively small numbers of days of DU and small number of cases. The result is low statistical precision and very wide CIs.

Female–Male Contrasts

Figure 3 depicts estimated female–male differences for AD 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_{50}. Conditional on the same rate of recent alcohol use, the estimated PD_{50} 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_{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 PD_{50} 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 number of newly incident users is substantial. For the PPR, one explanation for the 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 extramedical PPR use.

DISCUSSION

By turning to functional analysis of Hill equations, we offer a novel approach not previously seen in DD 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 extramedical 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 DU, 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 (ie, when no distinctions are drawn between those who just started to self-administer *versus* those who have been using the drug for more than 1 year). For cocaine, our P_{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% to 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, aggregating ever-users with cases found during two separate assessments. 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 and Anthony (2002) indicated that as many as 12%–13% of drinkers develop AD within the first 10 years after onset of drinking. The Lopez-Quintero study (Lopez-Quintero *et al*, 2011) estimated that 22%–23% of ever-drinkers had developed AD, with cases identified via two separate assessments. 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 (ie, 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%, once baseline and follow-up cases were counted (Lopez-Quintero *et al*, 2011). Estimates from Chen *et al* (2005) suggest that 4% 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 $P_{\min}$ of 2%, and for daily cannabis users the estimated $P_{\max}$ is 36%, irrespective of age.

This study's $P_{\min}$ and $P_{\max}$ values for newly incident extramedical users of prescription pain relievers are 4% and 38%, respectively. 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 OD 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 $P_{\min}$ and $P_{\max}$ for newly incident PPR users is 15.

As for female–male differences, there is a relatively limited epidemiology trace of published estimates (eg, 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 (eg, see Wagner and Anthony, 2007). For alcohol, there is some evidence of a more rapid transition from alcohol onset until AD for females (eg, see Hernandez-Avila *et al*, 2004; Mello, 1986; Zilberman *et al*, 2003). For cannabis, the female–male contrasts for these estimates are mixed (eg, 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 before assessment. The result is an 'apples *versus* oranges' comparison across drug subtypes. For tobacco, the focus is on dependence in the past month, ie, on the 30 days before assessment. For all other drugs, the focus is on dependence in past year, ie, on the 12 months before assessment. There is no way to recalibrate the NSDUH 30-day interval with its 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). In addition, the NSDUH team conducted detailed method studies on these topics (http://www.samhsa.gov/data/sites/default/files/2k6ReliabilityP/2k6ReliabilityP.pdf). These methodological studies do not indicate perfection in the assessments, but the findings from the methods research have been generally supportive.

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 a few 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, making us wonder whether large-sample longitudinal study estimates will ever 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 (ie, 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 our knowledge, no research project has attempted this intensity of measurement for fine-grained

time-sequenced drug experiences. It remains on an agenda for potential future research.

It is also important to remember that this study's estimates are based on 'DD' assessments based on pre-DSM-5 criteria. Large-sample epidemiological studies, often planned years in advance of the actual field assessments, tend to adapt slowly to changes in diagnostic criteria advocated by clinical psychiatrists. NSDUH is no exception. To date, the NSDUH DU disorder assessments are based on pre-DSM-5 diagnostic criteria (American Psychiatric Association, 2013). Once NSDUH assessments are adapted to DSM-5 criteria, it should become possible to repeat analyses of this type and to produce estimates for DSM-5 DU disorders in aggregate, and for 'DSM-5 drug addiction' as a separable category.

Conceptualizing extensions beyond this initial study, we propose that one way to accelerate this line of research is to formalize a multivariate vector of DD 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 DD 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 with those who qualify as diagnosable cases of DD. 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 extramedical use (eg, 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 (eg, manifestations of withdrawal on 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 (eg, 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 before the 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' (Lopez-Quintero and Anthony, 2015), as manifest in general NIH policy 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 DU.

FUNDING AND DISCLOSURE

In addition to peer review assignments mentioned in his financial disclosure statement, Dr Anthony recently completed a grant application review for the American Association for the Advancement of Science, for which he received a small honorarium compensation. The authors declare no conflict of interest.

ACKNOWLEDGMENTS

The work of OAV was supported by the National Institute on Drug Abuse (T32DA021129) and JCA's NIDA Senior Scientist and Mentorship Award (K05DA015799), and by Michigan State University. Author OAV declares that, except for income received from the primary university employer, no financial support or compensation has been received from any individual or corporate entity over the past thee years for research or professional service and there are no personal financial holdings that could be perceived as constituting a potential conflict of interest. The authors declare that over the past three years JCA has received compensation as extramural research reviewer, workshop speaker, consultant, or independent contractor from the National Institutes of Health (including Ole Consulting Group as NIH subcontractor), University of Miami (Florida, USA), Charles University (Prague, CZ), Korea University (Seoul, Korea), University of Ibadan (Ibadan, Nigeria), the Society for Epidemiologic Research (Clearfield, Utah), and the American Association for the Advancement of Science, and there are no personal financial holdings that could be perceived as constituting a potential conflict of interest.

REFERENCES

Ahmed SH, Koob GF (2005). Transition to drug addiction: a negative reinforcement model based on an allostatic decrease in reward function. *Psychopharmacology (Berl)* **180**: 473–490.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association: Washington, DC. Available at <http://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596>.

Ankomah P, Levin BR (2012). Two-drug antimicrobial chemotherapy: a mathematical model and experiments with Mycobacterium marinum. *PLoS Pathog* **8**: e1002487.

Anthony JC (2010). Novel phenotype issues raised in cross-national epidemiological research on drug dependence. *Ann N Y Acad Sci* **1187**: 353–369.

Anthony JC, Warner LA, Kessler RC (1994). Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: basic findings from the National Comorbidity Survey. *Exp Clin Psychopharmacol* **2**: 244–268.

Apelberg BJ, Corey CG, Hoffman AC, Schroeder MJ, Husten CG, Caraballo RS *et al* (2014). Symptoms of tobacco dependence among middle and high school tobacco users. *Am J Prev Med* **47**: S4–S14.

Bagnardi V, Rota M, Botteri E, Tramacere I, Islami F, Fedirko V *et al* (2014). Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis. *Br J Cancer* **12**: 580–593.

Bagnardi V, Zambon A, Quatto P, Corrao G (2004). Flexible meta-regression functions for modeling aggregate dose-response data, with an application to alcohol and mortality. *Am J Epidemiol* **159**: 1077–1086.

Becker WC, Sullivan LE, Tetrault JM, Desai RA, Fiellin DA (2008). Non-medical use, abuse and dependence on prescription opioids among U.S. adults: psychiatric, medical and substance use correlates. *Drug Alcohol Depend* **94**: 38–47.

Berlin JA, Longnecker MP, Greenland S (1993). Meta-analysis of epidemiologic dose-response data. *Epidemiology* **4**: 218–228.

Black JW, Leff P (1983). Operational models of pharmacological agonism. *Proc R Soc Lond B Biol Sci* **220**: 141–162.

Bobzean SAM, DeNobrega AK, Perrotti LI (2014). Sex differences in the neurobiology of drug addiction. *Exp Neurol* **259**: 64–74.

Brain P, Cousens R (1989). An equation to describe dose responses where there is stimulation of growth at low doses. *Weed Res* **29**: 93–96.

Chen C-Y, O'Brien MS, Anthony JC (2005). Who becomes cannabis dependent soon after onset of use? Epidemiological evidence from the United States: 2000-2001. *Drug Alcohol Depend* **79**: 11–22.

Chen K, Kandel D (2002). Relationship between extent of cocaine use and dependence among adolescents and adults in the United States. *Drug Alcohol Depend* **68**: 65–85.

Chen K, Kandel DB, Davies M (1997). Relationships between frequency and quantity of marijuana use and last year proxy dependence among adolescents and adults in the United States. *Drug Alcohol Depend* **46**: 53–67.

Del Boca FK, Darkes J (2003). The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addict Abingdon Engl* **98**(Suppl 2): 1–12.

Engel K, Sasaki T, Wang Q, Kuriyan J (2013). A highly efficient peptide substrate for EGFR activates the kinase by inducing aggregation. *Biochem J* **453**: 337–344.

Esser MB, Hedden SL, Kanny D, Brewer RD, Gfroerer JC, Naimi TS (2014). Prevalence of alcohol dependence among US adult drinkers, 2009-2011. *Prev Chronic Dis* **11**: 1–11.

Frank SA (2013). Input-output relations in biological systems: measurement, information and the Hill equation. *Biol Direct* **8**: 31.

Heeringa SG, West BT, Berglund PA (2010). *Applied Survey Data Analysis*. Chapman and Hall/CRC: Boca Raton, FL.

Hernandez-Avila CA, Rounsaville BJ, Kranzler HR (2004). Opioid-, cannabis- and alcohol-dependent women show more rapid progression to substance abuse treatment. *Drug Alcohol Depend* **74**: 265–272.

Kandel ER, Kandel DB (2014). A molecular basis for nicotine as a gateway drug. *N Engl J Med* **371**: 932–943.

Kasperski SJ, Vincent KB, Caldeira KM, Garnier-Dykstra LM, O'Grady KE, Arria AM (2011). College students' use of cocaine: results from a longitudinal study. *Addict Behav* **36**: 408–411.

Lopez-Quintero C, Anthony JC (2015). Drug use disorders in the polydrug context: new epidemiological evidence from a food-borne outbreak approach. *Ann NY Acad Sci* (in press).

Lopez-Quintero C, de los Cobos JP, Hasin DS, Okuda M, Wang S, Grant BF *et al* (2011). Probability and predictors of transition from first use to dependence on nicotine, alcohol, cannabis, and cocaine: results of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *Drug Alcohol Depend* **115**: 120–130.

Martins SS, Ghandour LA, Chilcoat HD (2007). Profile of dependence symptoms among extramedical opioid analgesic users. *Addict Behav* **32**: 2003–2019.

Mello NK (1986). Drug use patterns and premenstrual dystphoria. *Women Drugs New Era Res* 31–49.

Novak SP, Herman-Stahl M, Flannery B, Zimmerman M (2009). Physical pain, common psychiatric and substance use disorders, and the non-medical use of prescription analgesics in the United States. *Drug Alcohol Depend* **100**: 63–70.

Prins NH, Briejer MR, Van Bergen PJE, Akkermans LMA, Schuurkes JAJ (1999). Evidence for 5-HT7 receptors mediating relaxation of human colonic circular smooth muscle. *Br J Pharmacol* **128**: 849–852.

Reboussin BA, Anthony JC (2006). Is there epidemiological evidence to support the idea that a cocaine dependence syndrome emerges soon after onset of cocaine use? *Neuropsychopharmacology* **31**: 2055–2064.

Regoes RR, Wiuff C, Zappala RM, Garner KN, Baquero F, Levin BR (2004). Pharmacodynamic functions: a multiparameter approach to the design of antibiotic treatment regimens. *Antimicrob Agents Chemother* **48**: 3670–3676.

Ridenour TA, Lanza ST, Donny EC, Clark DB (2006). Different lengths of times for progressions in adolescent substance involvement. *Addict Behav* **31**: 962–983.

Ritz C, Streibig JC (2008). *Nonlinear Regression with R*. Springer: New York, London.

Royston P (2014). A smooth covariate rank transformation for use in regression models with a sigmoid dose-response function. *Stata J* **14**: 329–341.

Royston P, Sauerbrei W (2007). Multivariable modeling with cubic regression splines: a principled approach. *State J* **7**: 45–70.

Ryan H, Trosclair A, Gfroerer J (2012). Adult current smoking: differences in definitions and prevalence estimates–NHIS and NSDUH, 2008. *J Environ Public Health* (doi:10.1155/2012/918368; e-pub ahead of print 2 May 2012).

Sartor CE, Kranzler HR, Gelernter J (2014). Rate of progression from first use to dependence on cocaine or opioids: a cross-substance examination of associated demographic, psychiatric, and childhood risk factors. *Addict Behav* **39**: 473–479.

Seedall RB, Anthony JC (2013). Risk estimates for starting tobacco, alcohol, and other drug use in the United States: male-female differences and the possibility that 'limiting time with friends' is protective. *Drug Alcohol Depend* **133**: 751–753.

Seefeldt SS, Jensen JE, Fuerst EP (1995). Log-logistic analysis of herbicide dose-response relationships. *Weed Technol* **9**: 218–227.

Vaish AK, Folsom RE, Spagnola K, Sathe N, Hughes A (2013). An empirical study to evaluate the performance of synthetic estimates of substance use in the National Survey on Drug Use and Health. *Surv Res Methods Sect* 2538–2547.

Vignali C, Stramesi C, Vecchio M, Groppi A (2012). Hair testing and self-report of cocaine use. *Forensic Sci Int* **215**: 77–80.

Vsevolozhskaya OA, Anthony JC (2014). Confidence interval estimation in R-DAS. *Drug Alcohol Depend* **143**: 95–104.

Wagner FA, Anthony JC (2002). From first drug use to drug dependence; developmental periods of risk for dependence upon marijuana, cocaine, and alcohol. *Neuropsychopharmacology* **26**: 479–488.

Wagner FA, Anthony JC (2007). Male-female differences in the risk of progression from first use to dependence upon cannabis, cocaine, and alcohol. *Drug Alcohol Depend* **86**: 191–198.

Wetherington CL (2007). Sex-gender differences in drug abuse: a shift in the burden of proof? *Exp Clin Psychopharmacol* **15**: 411–417.

Wilhelm CJ, Hashimoto JG, Roberts ML, Sonmez MK, Wiren KM (2014). Understanding the addiction cycle: a complex biology with distinct contributions of genotype *vs* sex at each stage. *Neuroscience* **279**: 168–186.

Zernig G, Ahmed SH, Cardinal RN, Morgan D, Acquas E, Foltin RW *et al* (2007). Explaining the escalation of drug use in substance dependence: models and appropriate animal laboratory tests. *Pharmacology* **80**: 65–119.

Zilberman M, Tavares H, el-Guebaly N (2003). Gender similarities and differences: the prevalence and course of alcohol- and other substance-related disorders. *J Addict Dis* **22**: 61–74.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (http://www.nature.com/npp)