Published in final edited form as:

Ann N Y Acad Sci. 2010 February ; 1187: 353–369. doi:10.1111/j.1749-6632.2009.05419.x.

Novel Phenotype Issues Raised in Cross-National Epidemiological Research on Drug Dependence

James C. Anthony, Ph.D.

Abstract

Stage-transition models based on the American Diagnostic and Statistical Manual (DSM) generally are applied in epidemiology and genetics research on drug dependence syndromes associated with cannabis, cocaine, and other internationally regulated drugs (IRD). Difficulties with DSM stage-transition models have surfaced during cross-national research intended to provide a truly global perspective, such as the work of the World Mental Health Surveys (WMHS) Consortium. Alternative simpler dependence-related phenotypes are possible, including population-level count process models for steps early and before coalescence of clinical features into a coherent syndrome (e.g., zero-inflated Poisson regression). Selected findings are reviewed, based on ZIP modeling of alcohol, tobacco, and IRD count processes, with an illustration that may stimulate new research on genetic susceptibility traits. The annual National Surveys on Drug Use and Health can be readily modified for this purpose, along the lines of a truly anonymous research approach that can help make NSDUH-type cross-national epidemiological surveys more useful in the context of subsequent genome wide association (GWAS) research and post-GWAS investigations with a truly global health perspective.

Keywords

alcohol; tobacco; dependence; epidemiology; phenotype

1. Introduction

Writing this review my original intent was to provoke new action via radical thinking about new candidate phenotypes for research at the intersection of molecular genetics and epidemiological research on drug dependence (hereinafter, ‘genetic epidemiology of drug dependence’), as might be put to work in future genome wide association studies (GWAS), with large and rigorously drawn epidemiological samples of community residents in multiple countries. By ‘radical,’ I mean only that we will work back to the original roots of the ‘phenotype’ concept. In the process, some new phenotypes can be considered, which might prove to be useful in the genetic epidemiology of drug dependence, particularly as we try to gain an initial multi-national and then a global perspective on this topic.

In some respects, the review has the quality of an editorial of a type written about ten years ago to advocate a major reorientation of scientific work on the genetics of bipolar disease and schizophrenia (1). What I am trying to accomplish is to reorient research on genetic epidemiology of drug dependence, and in the process I will make a very practical recommendation about how this work can be accelerated via a re-conceptualization of the

Correspondence: James C. (Jim) Anthony, Ph.D., Professor of Epidemiology, College of Human Medicine, Michigan State University, East Lansing, Michigan 48824, janthony@msu.edu, Tel: 517-353-8623 x100, Fax: 517-432-1130.

Conflicts of interest: The author declares no conflicts of interest.

phenotypes under study. I provide a practical example and describe experiences in a recent pilot study that was intended to set in motion a future elaboration of large-sample epidemiological field studies that now are not being made informative with respect to the genetic epidemiology of drug dependence.

One facet of what I shall recommend involves a return to practices of the past. Another facet involves new points of departure. In specific, the acceleration of progress I have in mind may require somewhat of a departure from the drug dependence or 'addiction' phenotypes as they presently are framed in contemporary American psychiatry, and which may not be as useful outside of America as they are in the American context.

What I will share may come as a surprise to readers who know my work as an epidemiologist who mainly has tried to extend trail-blazing paths laid by the late Professor Lee Nelkens Robins, together with Professor John E. Helzer, with respect to a focus on the drug dependence concept derived from contemporary diagnostic criteria and case definitions as originally framed in the American Psychiatric Association's Diagnostic and Statistical Manual, Third Edition (DSM, DSM-III), and in the later DSM-IV. In the end, I shall recommend a deliberate departure from that research approach in order to accelerate progress in drug dependence epidemiology, and to promote creation of parallel tracks of research. That is, I judge that for a time we can accelerate the progress of these scientific fields by working along non-intersecting tracks that run alongside the still- evolving DSM diagnostic traditions and guide the clinical practice of psychiatry.

My proposal is that the greatest acceleration might be achieved via parallel track research over the next 5-10 years, with an occasional 'crosswalk' across a bridge first built orthogonally, until there is a sufficient accumulation of evidence about non-parallelism of the tracks. I will provide a conceptual model, illustrated with a descriptive figure, in order to make clear that we must begin with non-parallel tracks because in virtually all observational research of our time the field has ignored a reciprocity (sometimes erroneously termed 'reverse causation') when prior crosswalks have been created.

Before presentation of the figure, a note about Professor Lee Nelkens Robins may be in order in that she passed away no more than a few months ago, and her work may not be well-known in some sectors of this field. Lee was a trailblazer in 20th century psychiatric epidemiology generally and in drug dependence epidemiology specifically. Her PhD sociology studies with Talcott Parsons at Harvard University occurred when Parsons was attempting to move structural functionalism in the direction of action theory and the AGIL framework of institutional necessities for successful society formation and development. [AGIL is an acronym for adaptation (A) to environmental conditions, consensus setting with respect to goal attainment (G), harmonization and integration (I) of the society as in adoption of common values, norms, and language, and successful development of latency (L) mechanisms such as bonds to family and school through which values, norms, and language are passed from generation to generation.] For Lee's collaborators who were aware of Parsons' AGIL framework, it was obvious to see her attempts to harness that framework in the creation of the small social groups through which successful scientific progress might be made.

From the mid-1970s forward, she taught many of us a conceptual model for drug dependence that I have characterized as a stage-transition model, and have put to good use in the study of tobacco dependence and related processes, often harnessing latent class and transition approaches (e.g., 2-5). Lee taught that the first stage in the drug dependence process was having a chance or opportunity to try a drug. Borrowing from Wade Hampton Frost's concept of 'exposure opportunity,' first expressed some 80 years ago in a theory of

infectious disease epidemiology, my own research group has worked up various measurements of that first chance to try a drug. We now have evidence of a fairly general proposition that a male excess in occurrence of drug dependence often has its origins in a male excess in the timing of the first drug exposure opportunities. In most countries in which we have conducted our studies, age by age, males are more likely to have had a chance to try illegal drugs, as compared to females in the same birth cohorts. Once the chance to try occurs, the next stage-transition is actual use of the drug. Here, the male-female variation is much attenuated in general; females seem to be just as likely as males to use the drug, once the chance to try has occurred. Then, once drug use occurs, some drug users develop a drug dependence syndrome while others do not. Sometimes there is a male-associated excess conditional risk of becoming dependent once use starts; sometimes not. For example, we have found male-female differences in the risk of becoming cannabis dependent soon after onset of cannabis smoking, less pronounced male-female differences for alcohol, and even smaller male-female differences for cocaine (2-10). As a side note, one might contemplate epidemiological research on the later stage-transitions (e.g., from dependence into recovery), but our research group always has been concerned about attrition biases in the large-sample epidemiological research context once the drug dependence process starts – given excess mortality risk attributable to drug dependence, as well as potential under-representation of heavy drug users in community sample surveys (11).

The first formal epidemiological work on the stage-transition model from drug use to dependence was completed by John Helzer during a followup study of male Vietnam era veterans, for which Lee Robins was Principal Investigator (12). No complex latent transition modeling was involved in that research. Relatively simple contingency table analyses proved to be enough to convey that (a) most veterans who had used opioid drugs during Vietnam era service had not developed sustained opioid ‘addiction’ problems that complicated their lives back at home, and (b) some characteristics were predictors of the post-service ‘addiction’ outcome while other characteristics were not predictive at all.

Seeking examples from the published literature for this review article, I came across some interesting elaborations of the stage-transition model that didn't make much sense to me. To illustrate, consider a simple plot of drug dependence prevalence estimates on the y-axis of an x-y graph, with the count of the number of days or occasions of drug use on the x-axis. This plot can be envisioned, and with appropriate estimates and data, it can be plotted. But it is valid to draw this plot only if the relationships linking the counts of drug experiences and the occurrence of the drug dependence process are in conformity with the acyclic assumption of standard dose-response modeling.

I believe that behavioral pharmacologist pioneer Joseph Brady was the first to recognize a violation of the acyclic assumption in relation to the process of starting to use a drug and then becoming dependent upon it, but his regrettably neglected article on the topic is buried within a NIDA research monograph that has not been cited frequently (13). In Figure 1, I have re-drawn Brady's earlier and more artistic torus-like structure of feedback loops that link repetitions of drug self-administration experiences with the formation of clinical features of drug dependence. In this rendition, I have tried to give a very clear depiction of a violation of the acyclicity assumption, with the drug dependence features feeding back and influencing the repetition of drug self-administration experiences. Readers familiar with the work of Koob and LeMoal (14), among others, will be familiar with this type of feedback loop and re-setting of set points in the processes of neuroadaptation and repetitive drug self-administration.

Figure 1 shows the onset of an individual's drug use at 1. At Step 2, there is a piling up of occasions or days of drug use after first use. Thereafter, at Step 3, tangible clinical features

of drug dependence begin to show up (e.g., subjectively felt tolerance, craving). At some point before or during the formation or coalescence of these individual clinical features into a drug dependence syndrome, there is a feedback loop such that the syndrome formation process begins to drive up the count of drug-taking occasions and the associated rate of drug use per unit time, just as the accumulating count of drug-taking occasions drives forward the drug dependence process.

With a reciprocity or feedback loop of this type, the acyclicity assumption of standard dose-response modeling is violated. It becomes difficult to make sense of a plot that expresses the prevalence of drug dependence as a function of the count of drug-taking experiences as if this were some standard type of acyclic 'dose-response' relationship. Here, the response is allowed to influence and drive up the size of the 'dose' in a clear violation of assumptions for standard statistical modeling of dose-response relationships.

Advanced models exist for use when the acyclicity assumption is violated, but the data for these models necessarily must be longitudinal in character. In the context of the multi-national or global GWAS investigations mentioned in the first paragraph of this article, to gather such longitudinal data on the scale of epidemiological sampling for a GWAS investigation would drive up the cost of any multi-regional global drug dependence epidemiology research program, not to mention a need for future replications with equally sized samples and other refinements to satisfy the GWAS critics.

A partial cross-sectional study solution might be derived by yoking an appropriate model for count or rate responses with an appropriate model for multivariate response profile of clinical features, allowing the interdependency of these count and multivariate response constructs to be estimated in a correlation that is agnostic with respect to direction of influence (e.g., drawn in a Directed Acyclic Graph with a double-headed curved arrow). The cross-sectional sample to model these relationships within the first month after onset of drug use would have to be quite large in order to include a sufficient number of earlier-onset users of recent onset vintage. Estimation of the parameters of separate models, month by month, for each month of elapsed time after onset of drug use could foster development of a cross-sectional mosaic of the time course of the drug dependence process, expressed as a function of genes and environmental conditions and processes interacting. The result would be a useful development, much needed to derive starting estimates for planning of prospective and longitudinal research for more definitive evidence on these interrelationships.

Returning to Figure 1's steps 1 and 2, it is possible to see that the drug dependence process actually begins with a count process that predates the formation of the drug dependence syndrome. This count process might be measured as the number of self-administrations per interval of time under study, or the number of occasions of use or number of days of use during that interval.

Recognition of these developments that antedate the clinical features of the drug dependence process prompted our research group to turn its attention to count process models that might be used in research on the very early stages of drug dependence. It is fortunate that in the past 20 years there have been remarkable advances in the statistical methods for addressing questions about count process parameters of considerable interest in the study of the process of becoming drug dependent. In a later section of the article, a count process model is described in more detail, with an illustration that forecasts what might be done in large multi-national epidemiological research to complement current approaches based on the stage-transition tradition pioneered by the Robins research group, including our own research in genetic epidemiology with community and other samples (e.g., see 15). As

compared with the genetic epidemiological studies of dependence on internationally regulated drugs such as cannabis and cocaine, there has been a much greater diversity of phenotypes under study in genetics research on the dependence syndromes involving tobacco and alcohol. To illustrate, one might consider the broad range of phenotypes outlined in a recent National Cancer Institute (NCI) tobacco control monograph that can be downloaded from a web site (16). Of course, there is an advantage to the Robins tradition of orienting the stage-transitions studied in epidemiology to the same diagnostic constructs for drug dependence that are being used in the clinical practice of psychiatry and allied professions responsible for treating patients with these conditions. The resulting epidemiological estimates have clinical implications that are not too difficult to communicate to clinicians who are treating or doing research with these patients. Epidemiologists' use of the DSM-type diagnostic constructs also seems to serve well in the introductions and background sections of pre-clinical research reports, helping to substantiate the public health significance of the drug dependence syndromes.

In the rest of this article, I will review two lines of research that may help motivate a departure from this time-honored stage-transition tradition. One of the lines of research is based upon the work of the World Mental Health (WMH) Surveys Consortium, for which I have some responsibility as an affiliated Principal Investigator. The other line of research is based upon the work of the Office of Applied Studies of the Substance Abuse and Mental Health Services Administration (SAMHSA), which directs the National Surveys on Drug Use and Health (NSDUH) and then makes NSDUH public use datasets available to non-governmental research groups such as the one for which I am responsible.

The WMH Surveys Consortium is important in this context because it is taking the Robins tradition of standardized diagnostic assessment of drug dependence and related conditions into multiple countries and all regions of the world, with a forward plan for research on the genetic epidemiology of drug dependence as well. The NSDUH is important because its sample size produces enough newly incident cases of drug dependence and related phenotypes to make feasible future genome wide association studies (GWAS), with multiple replications of the type required to distinguish falsely positive leads from candidate genes that deserve more attention and work-up in efforts to understand pathophysiological pathways and molecular mechanisms with the most public health importance.

Nevertheless, in these two lines of research, we see reason and opportunity to broaden the ranges of candidate phenotypes under study, including a return to the original concept of a candidate phenotype as a population characteristic and not just an individual-level characteristic of each organism. Due to the emphasis upon phenotype as an individual-level characteristic, many genetic epidemiologists working in the field of drug dependence have forgotten or never learned that Wilhelm Johanssen originally coined and defined the term 'phenotype' with a population perspective that complements a perspective on phenotype as a characteristic of an individual. This population perspective motivates thinking about the 'phenotype' in terms of what we might call the moments of the statistical distribution from a population. The foundation of individual-level research approaches in modern genetic epidemiology have grown out of post-Johanssen advances in cytology and work at the intersection of molecular biochemistry and genetics (17-18). It now may be useful to think of some population-level candidate phenotypes for new work at the intersection of molecular genetics and epidemiological research on drug dependence. This thinking is prompted in part by some recent observations from the World Mental Health Surveys, which will be discussed next. The thinking also is prompted by a consideration of potential alternatives as might be applied in future genetic epidemiology studies that can exploit the large sample replications of the NSDUH with some minor elaborations, which will be reviewed after coverage of the WMH Surveys Consortium observations.

2. Selected Observations from the World Mental Health Surveys Consortium

Several years ago, our research group was invited to join the World Mental Health Surveys Consortium and to develop a new research plan for analyses of field survey data on drug dependence from what now are the more than 20 WMHS sites in all of the World Health Organization global regions in this important effort to build a multi-national and global perspective on psychiatric disturbances (e.g., see 19-21). Necessarily, true to its character as epidemiological surveillance research, these field investigations are designed to yield timely, practical, and cost-efficient results as might guide policy and public health planning efforts in each participating country. As such, the WMHS borrowed the assumptional framework of DSM diagnosis, and the WMHS diagnostic assessment plan was designed in accord with the DSM-IV, with some acknowledgment of the ICD tradition (22). That is, the WMHS drug dependence measurements depend heavily upon the DSM diagnostic assumption that drug dependence is a clinically significant condition, with at least some manifestations of impairments in social and occupational functioning, legal difficulties, or recurrent drug-related hazard-laden or socially maladaptive drug use (e.g., see 20,23-24). Otherwise, following DSM-IV criteria, the WMHS standardized field survey assessment of drug dependence is based upon a familiar multi-item approach with items that tap neuroadaptive changes secondary to repetitive drug-taking (e.g., subjectively felt tolerance, withdrawal), drug-related disturbances of the mental life (e.g., obsession-like craving or strong desire to take the drug), and drug-related disturbances of behavior (e.g., compulsion-like failure to stop or cut down on drug use even when there is an intent to do so). The standardized survey research items to tap these facets of drug dependence originally were written in English, in an adaptation of the original Robins' approach for the National Institute of Mental Health Diagnostic Interview Schedule and the World Health Organization Composite International Diagnostic Interview (CIDI), with later translation, back-translation, and harmonization to create the interview schedules for the other language groups for this multi-national research (e.g., Arabic, Portuguese, Hindi). Necessarily, there was an assumption that these items would perform equally well across national and language boundaries in their measurement of the drug dependence syndromes within the field survey context – at least as well as they had performed in the original CIDI trials based mixed clinic-community samples of drug users (e.g., see 25). To some extent, this assumption is being tested in the latent structure analyses described below, within this article.

In each country or site under study, one of the WMHS Principal Investigators has adapted a common research approach to secure institutional review board approval. The approach is one that involves designation of a source population for this epidemiological research, construction of a sampling frame for multi-stage area probability sampling of household dwelling units (DU) and then probability sampling of designated respondents within each sampled DU, followed by the standardized field survey assessments of a broad range of psychiatric and behavioral disturbances and related characteristics. In some sites, the assessment involves an interviewer reading from the standardized interview schedule and marks the DR's responses using a paper-and-pencil interview approach; in other sites, the DR completes a computer-assisted self-interview on a laptop computer after an introduction by the field staff member. At most sites, 70% or more of the sampled DR agreed to participate, and the resulting sample size at each site has ranged from just over 2300 adults to almost 13,000 adults. For example, in the United States, the participation level was 71% and the resulting sample size was 5,692. Detailed methodological descriptions of the WMH Surveys have been published and are widely available for readers interested in these details (19-22).

For the latent structure analyses, we are using a maximum likelihood approach for an exploratory factor analysis (EFA) to check model fit for varying numbers of underlying dimensions and to estimate item response theory (IRT) parameters under the often-confirmed unidimensional model, with appropriate weighting of the survey data to the sampling probabilities into account. MPlus software, version 5, is being used in these analyses (26).

The first model specified for this challenge is one that is consistent with the DSM-III-R and DSM-IV assumption that there is a list of exchangeable clinical features of drug dependence – exchangeable in the polythetic sense that a diagnostician might count up the number of clinical features listed under each diagnostic criterion, no matter which combination of clinical features is assembled in the count. Then, provided that the resulting count of clinical features exceeds a threshold value, the diagnosis of drug dependence is advanced, subject to other criteria (e.g., the general ‘clinical significance’ criterion stipulated in the introductory paragraph that precedes the DSM listing of the exchangeable clinical features). In this DSM diagnosis assumption, we have an implicit assertion that each listed criterion or clinical feature is equal in its capacity to discriminate the positioning of the examinee on an underlying single dimension of drug dependence, and that drug users who are positioned above the cut-off threshold on that dimension should be looked into more thoroughly with respect to the other diagnostic criteria. (An alternative to the equal-weighting approach is one in which some clinical features are given more weight than others, with weights derived on the basis of theory or on the basis of evidence such as the values of IRT discrimination parameters of the type being estimated with the WMHS latent structure approach.)

Figure 2 illustrates one set of de-identified WMHS estimates that result from these EFA latent structure analyses. The plotting approach is one we use to summarize results from latent structure analyses of this type. The identity of these specific WMHS sites is not important for present purposes; the individual site Principal Investigators are in the process of publishing this work for the first time. Here, the de-identified patterns are used to provide readers with a clear impression of variability across sites in the resulting parameter estimates, implying that the clinical features (as manifest in the WMH survey assessments) perhaps should not always be weighted equally in the diagnostic assessment.

In the form of a ‘radar’ plot, Figure 2 shows for each of the six de-identified study samples a set of estimates of how well the DSM list of drug dependence clinical features helps to discriminate the underlying single dimension of drug dependence. This modeling is based on a stipulation that there is just one underlying dimension that captures dependence upon internationally regulated psychoactive drugs such as cannabis, which is consistent with the DSM-III-R and DSM-IV specification to impose a cut-off threshold on a count of the clinical features that tap each diagnostic criterion under the drug dependence rubric. For one of the sites (row 1, column 2, labeled X12), with respect to the discrimination parameter, the exchangeability of the clinical features or criteria listed in the DSM-IV is manifest in estimates that do not differ appreciably, feature-by-feature, as one looks around the plot to each projecting axis, where each axis represents a different one of the listed drug dependence criteria or clinical features as measured in the WMHS study samples. Two of the measured criteria did not serve well, as represented by the ‘pie slices’ in that same X12 plot at the 12 o'clock position and roughly at the 3-4 o'clock position. Otherwise, as indicated by the shading, the estimated discrimination parameters were not appreciably different from one another.

In contrast, within Figure 2, there is some degree of similarity in the ‘radar’ plots for the sites labeled as X12 and X21, but the argument that the clinical features are ‘exchangeable’ is more difficult to make in the X21 results. For example, in X21, the shading shows larger

discrimination estimates for the clinical features depicted in the 5-11 o'clock positions, with smaller discrimination estimates for clinical features depicted in the 12-4 o'clock positions.

Figure 2 also gives much different configural pattern that is seen for the sites labeled X13 and X23 (i.e., different from what is seen for sites X12 and X21). In specific, the X13 and X23 discrimination parameter estimates follow similar patterns, with low discrimination found for clinical features depicted at the 3-4 o'clock and 7 o'clock positions, but with an otherwise general similarity for the other estimates.

Interested readers are referred to the primary WMHS articles on these analyses, now in preparation, that will convey these findings in more detail. Nonetheless, preliminary plots of this type illustrate one of the challenges to the DSM diagnosis assumption in drug dependence epidemiology, if we are to seek a multi-national or global perspective. It is quite clear that broad variation in these item-level performance characteristics of the DSM diagnostic assessment can be an impediment to their general use in cross-national GWAS on drug dependence. An alternative approaches must be specified. At this juncture, we are reviewing less complex new candidate phenotypes that can be used in complement with the CIDI diagnostic assessments in a line of genetic and other epidemiological field studies to investigate drug dependence processes.

This is an illustration of understandable measurement problems that now are being faced when the largely American DSM-IV concepts for drug dependence, with western European traditions as well, are taken into the non-American, and often non-western, social-cultural milieu of community-based epidemiological samples. Our proposed research approach for future multi-national research involves a re-orientation in relation to the count processes mentioned above, which predate the emergence of the clinical features of drug dependence we have been studying in the WMHS Consortium. As compared with the measurement of more abstract clinical concepts secondary to neuroadaptation, and drug-related disturbances of the mental life or behavior, the measurement of facets of the count processes can be made much less complex. Fairly straightforward survey items on the number of days of drug use in fixed intervals of time after onset of drug use or prior to assessment are required (e.g., 30 days intervals). As it happens, in most parts of the world, American or non-American, western or non-western, the number of days of drug use in a span of 30 days is a reasonably well-understood concept, not so difficult to measure. Some of the tobacco dependence phenotypes presented in the NCI monograph cited above are based upon survey or clinical assessments of this type; their application in genetic epidemiological research on cannabis, cocaine, and the other internationally regulated psychoactive drugs is not yet as widespread as the reliance upon stage-transition models and associated measurements.

3. Selected Observations from the National Surveys on Drug Use and Health

In the historical background of the present work is one of the accomplishments of Charles R. Schuster as Director of the National Institute on Drug Abuse, which was to issue a directive that set into motion the creation of public use datasets from NIDA's National Household Survey on Drug Abuse (NHSDA). This directive opened important possibilities for 20+ years of new epidemiological research by extra-governmental investigators, including our own extramurally funded university-based research group (e.g., 27). Later, governmental responsibility for the national surveys program was shifted to SAMHSA and the surveys have been re-named as the National Surveys on Drug Use and Health, but the Office of Applied Studies at SAMHSA has continued to supervise the creation of public use datasets based on the NSDUH field work each year.

The now regularized NSDUH research approach for the United States (US) is akin to the WMHS in many respects, but the samples are much larger. For example, there is multi-stage area probability sampling of non-institutional dwelling units (with homeless shelters and other group quarters now listed within the NSDUH sampling frame). Within each DU, there is probability sampling of one or more designated respondent-residents who is asked preliminary questions about himself or herself and the other occupants of the DU. The participation level for each year's survey has declined somewhat over the years, but in recent years it has been at respectable values of 65%-70%, and the resulting sample size for the government's research is 67,760 individuals age 12 years and older. The public use datafile released for non-governmental research groups is somewhat smaller and contains only 55,602 records after a sub-sampling step that is used to protect identities of participants and to thwart inadvertent or deliberate attempts to identify individual participants. There is a audio-enhanced computer-assisted self interview (ACASI) approach with standardized modules and multiple items on drug use and related topics, after an introduction to use of the computer keyboard and devices to record responses.

Consistent with its service as an epidemiological surveillance tool, most of NSDUH assessment modules now are concentrated on the quantity, frequency, and variability of each designated respondent's drug use history. Nonetheless, guided by Robins' work on drug dependence as part of the NIMH Epidemiologic Catchment Area studies of 1980-84, the NSDUH staff has introduced modules on recently active drug dependence clinical features. The standardized items on these individual clinical features of drug dependence are much like those written in the original English version of the WMHS interview schedule, but the time interval under study is not entire lifetime history of the individual who self-identifies as a drug user, but rather is focused upon drug dependence-associated experiences during the 12 months prior to the date of assessment. To permit study of time trends, these modules are part of the core assessments for the NSDUH that have been held constant for five or more years. There is concern that changing the NSDUH assessment (e.g., to introduce past history of drug dependence) might distort the study objectives with respect to these time trends, which require stability in the research approach.

As outlined in a series of studies published over the past decade, our research group generally re-approaches the NSDUH data analyses with a focus upon the newly incident or 'recent-onset' drug users in order to shed light on the earliest stages after onset of drug use, and upon the emergence of clinical features of drug dependence during the first 12-24 months after onset of use of each drug. This type of inquiry involves Step 3 depicted in Figure 1, where clinical features of drug dependence begin to emerge and sometimes to coalesce into a syndrome.

Figure 3 provides an illustration for one line of research of this type, in which the hypothesis under study was whether the clinical features of tobacco dependence might emerge more rapidly among early-onset tobacco smokers, age 11-17 years at the time of smoking onset, as compared to adults with smoking onset after age 17 years, and with elapsed time from smoking onset to the date of assessment constrained to be a relatively short interval of no more than 24 months. As shown in Figure 3, when studied one by one, there is little age-of-onset-related variation in the risk of experiencing each clinical feature within the first 24 months after tobacco smoking onset; this conclusion has been confirmed with multivariate response regression modeling (28). Instead, it would seem that previously reported excess risk of tobacco dependence for early-onset smokers might be a function of elapsed time during the intervals between smoking onsets and the date of post-smoking assessment. The need to hold constant elapsed time since onset of drug use in research of this type originally was noted in relation to studies of illegal drug use and later risk of drug problems (29) and the application of this general principle in the study of early-onset tobacco smoking is not

new (28,30). [For readers interested in each of the clinical features listed along the x-axis of Figure 3, the legend for the abbreviations is as follows: G, Spend more time getting drugs; O, spend more time getting over the effects; U, unable to keep to limits on drug use; S, needed more drug to achieve same effect; L, using same amount but giving less effect; C, unable to cut down; E, continue to use despite emotional problems; P, continue to use despite physical problems; A, reduced activities due to drug use; W, withdrawal symptoms.]

In more recent work, we have shifted the focus of inquiry in the direction of Step 2 in Figure 1, and have studied the count process through which the first occasions of drug use might start to drive forward the drug dependence process before this count process becomes driven by the dependence process in the type of feedback loop discussed in this article's introduction. By envisioning and studying this early step in relation to a count process that does or does not move forward after first use of most psychoactive drugs, a statistically minded scientist may recognize a problem of zero-inflation relative to standard Poisson distributions – i.e., more heterogeneity than the statistical mean of the Poisson distribution would imply. This excess of zero counts of drug experiences after the initial drug self-administration can be used to help motivate two parallel lines of research on new phenotypes in epidemiological research on genetic and other determinants of the drug dependence process. The first line of research concerns the characteristics of individual drug users, and the conditions and processes that might account for variation in who takes a drug one time only, and then never again, which by itself deserves greater scrutiny as a phenotype of interest in relation to the internationally regulated psychoactive drugs, and also in relation to tobacco smoking research where there often has been a restriction to individuals who have smoked 100 cigarettes or more, not only in the phenotypical cases, but also in the controls (e.g., see 31). The other line of research concerns the statistical parameters of the count process, studied in a fashion that takes us back in the direction of Johanssen's original conceptualization of the phenotype with a population perspective and a complementary individual-level perspective (17-18).

In order to understand the population phenotype in mind, one must start by coming to grips with the epidemiological observation that a great many drug users never develop a drug dependence syndrome (e.g., see 28,32). Indeed, as noted above, a great many drug users try a drug once but then never try it again; this is seen especially for the internationally regulated psychoactive drugs. Note that during the first 30 days after drug use starts, including the first day of use, there is an absence of zeroes that should be seen in the Poisson distribution; it is a zero-truncated distribution because, by definition, all drug users who start to use in a given month have at least one occasion and one day of drug use during that month. Thereafter, in most subsequent months, there is an *excess* of zeroes in that for virtually all drugs and virtually all recent-onset drug users, the first month of use might include 2 or 3 more days of use, but thereafter, the zero day value predominates and there is no persistence of use. These are our zero-inflated Poisson distributions.

Just as there are regression models for truly conforming Poisson count distributions with no zero inflation, there also are regression models for zero-truncated and zero-inflated Poisson distributions. This is where the concept of a population phenotype comes back into the picture, in conjunction with the concept of an individual phenotype. Namely, the regression model for zero-truncated count distributions during the first month of use can be converted into a zero-inflated count distribution by bringing into play all of the individuals who never tried the drug after the first occasion of use, and for whom the value for the 30 day interval prior to assessment is zero days of use. In other words, these never-again users would be joined with the persisting (still-using) recent-onset users to create the count distribution for that 30 day interval. (Those with onsets of drug use in the distant past are not contributing information, even if they persist in drug use, because they are not members of the newly

incident or recent-onset inception cohort of the newest drug users in the sample, as in 28). This approach can be used to unify the statistical treatment of the first 30 days after onset of use and each subsequent 30 day interval in that all of the resulting count distributions will tend to follow the zero-inflated Poisson pattern because many people try a drug one or two times and never try again.

Moreover, in each month's aggregate of individuals with zero days of drug use, we can conceptualize two latent classes, with one class (U) consisting of 'Subsequent Potential Users' (or 'Not Always Non-Users') and the other class ($\sim$ U) consisting of 'Subsequent Always Non-Users' (i.e., subsequent to the first day of using the drug). These classes are latent in that they are observed indirectly, and it is not possible to sort individuals into the U-bin and the $\sim$ U bin, although it is possible to estimate the average probability of an individual being in the U-bin and in the $\sim$ U-bin. (The reason we do not observe this class membership directly is because some of those with zero days of use in one specific month may go on to use the drug again in a future month.) This is the population phenotype mentioned in the introduction for this review article; the average probability is a summary of a statistical distribution for the population under study, as Johanssen had in mind when he coined the term 'phenotype' (17-18).

As our research group has developed greater experience with the zero-inflated Poisson model, we have started to appreciate its potential for applications in the study of drug dependence processes, the count process that predates emergence of the first clinical features of drug dependence, and the count process that then starts to be driven by drug dependence. In specific, the zero-inflated Poisson (ZIP) regression model developed by Lambert can be refined so as to accommodate the excess of zeroes in the count distribution by changing the mean structure, with the probability of a zero value allowed to differ across population subgroups as indexed by covariates in the 'inflate' part of the ZIP regression model (33). The ZIP regression model involves a regression of membership in the two latent classes (the U-bin of 'Potential Users' versus the $\sim$ U-bin of "Always Non-Users") upon covariates that can be constructed to reflect subgroups of importance in the genetic epidemiology of drug dependence, perhaps based upon: (a) number of copies of an allele of substantive importance, (b) haplotype structure, (c) gene-gene combinations, (d) gene-protein combinations, (e) gene- or haplotype-environment combinations, and (f) environment-environment combinations. Strong positive associations with membership in the 'Always Non-Users' $\sim$ U class would direct attention toward possibly protected subgroups; strong inverse associations would direct attention toward possibly high risk subgroups (33-34). Again, what distinguishes this approach to formation of latent classes in contrast to the other latent class models that have become popularized is its reliance upon the statistical theory. Nonetheless, unlike the other latent class models discussed in Section (2) of this article (e.g., see 2-5), there are no survey-assessed items to tap membership in the latent classes; there is no direct assignment of individuals in terms of probable class membership. Probabilistic statements about subgroups of individuals can be made by inserting actual subgroup values into the 'inflate' part of the regression equation, but these are subgroup averages. It is not possible to point to a specific individual with a zero count and to declare with certainty whether that drug user is a member of the U-bin or the $\sim$ U-bin.

In contrast to the 'inflate' part of the ZIP regression model, the count part of the model expresses log differences in the count of days of use per unit time as a function of the individual-level or subgroup covariates. This log difference also can be a phenotype of interest, as is the antilogged expression, which is a rate ratio (e.g., number of days of drug use per the most recent 30 days prior to assessment, for each subgroup indexed by covariate values; 33-34).

Combining information from the ‘inflate’ and the ‘count’ equations in the ZIP regression model, one might forecast identification (e.g., via haplotype or haplotype-environment combination) a population subgroup that is highly likely to be over-represented in the U-bin of “Subsequent Potential Users,” rare to be seen in the ~U-bin of ‘Subsequently Always Never Users’, and with a very high rate of drug using days during Month $t+k$ after onset of drug use in baseline Month t . This subgroup, identified by the mean structure of the population phenotype in the form of the zero-inflated count distribution, would be one of considerable interest in this new line of research in relation to the genetic epidemiology of drug dependence.

One also might forecast similar identification of a population subgroup that is under-represented in the U-bin of “Subsequent Potential Users,” but that is no different from other subgroups with respect to rate of drug use once it has occurred. By thinking along these lines, we hypothesize that Asian-Americans might be a subgroup of this type in relation to the count of days of recent alcohol use. Our reasoning was that alcohol-related toxicity might prompt some Asian-American subgroups to drink alcohol one time and then never drink it again, yielding to their under-representation in subsequent months with respect to the “Subsequent Potential Users” latent class (e.g., by virtue of homozygosity with respect to the variant *ALDH2**2 gene allele, with an allowance that some Asian-Americans who are heterozygous might drink and some might not, as in 35). Nonetheless, conditional on the latent class outcome, to the extent that some members of the subgroup might end up in the ‘Potential Users’ class, if they persisted in drinking in months after the first month of drinking, their rate of drinking (days of drinking per 30 days prior to assessment) can be posited to be no different from the rate of drinking for other subgroups (e.g., a reference subgroup of non-Hispanic White Americans).

Borrowing information from the multiple years of the National Surveys on Drug Use and Health conducted since 2002, Brian Fairman and other members of our research group recently identified almost 13,500 individuals who had started to drink alcohol within 24 months prior to the date of assessment. We have stratified these newly incident and recent-onset drinkers by an approximation of the number of days that passed from the first drinking day to the most recent drinking day. For almost 6000 of these recent-onset drinkers, we can gauge that there was no more than about 90 days that passed between the first and the most recent day of drinking. For almost 1300, an estimated 91-180 days passed from the first day of drinking to the day of most recent drinking. Our general approach to the study of subgroup variation in the count process is one that follows the first drink in the direction of an increasing count of drinking days, with a sorting of the newly incident recent-onset drinkers into categories defined by (a) elapsed time since onset of drinking (here, always with focus on those who had started drinking within 24 months of the date of assessment), and (b) elapsed time from the first drinking day to the most recent drinking day (28). Within this framework, we now are applying the zero-inflated Poisson regression model to evaluate hypotheses about population subgroups that should or should not persist in drinking or that should have lower rates of drinking, conditional on persistence of drinking.

As hypothesized, with or without age and sex held constant, and as compared to the reference subgroup of non-Hispanic White American recent-onset drinkers, the Asian-American recent-onset drinkers were more likely to belong to the ~U latent class of “Subsequently Always a Never Drinker” (in the 30 day intervals under study; $p < 0.05$). Nonetheless, conditional on that latent class outcome, the Asian-Americans who drank did so at the same rate as the non-Hispanic White drinkers ($p > 0.05$).

In this extension of our research on the comparative epidemiology of the earliest stages of drug involvement (10,32), we find no evidentiary support for the idea that the ZIP regression

count process parameters differ for Asian-American recent-onset drinkers as compared to non-Hispanic White recent-onset drinkers for any other psychoactive drug we have studied. Alcoholic beverages are the exception, for which there is biological plausibility for non-persistence of drinking in relation to the previously mentioned genetics-related alcohol toxicity (35). Table 1 summarizes our findings on four psychoactive drug compounds studied in this fashion, based on recent work by the group, which should appear in print in the coming months (Meyer et al., under review; Barondess et al., under review; Mainampally et al., under review; Anthony et al., under review).

(4) Proposed Future Elaboration of the NSDUH and WMHS beyond 2010

In light of these findings from the World Mental Health Surveys and the National Surveys on Drug Use and Health, it may now be useful to return to the original proposition that a departure from the stage-transition models may be necessary if we are to develop a more global perspective on the genetic and environmental determinants of the process that leads from the earliest stages of drug involvement toward risk of becoming drug dependent. Initial optimism that this global perspective might be developed with a single-minded focus upon the stage-transition model developed by pioneers Robins and Helzer is dampened to some degree by the WMHS observations from latent structure analyses that indicate that the DSM-IV concepts for the drug dependence syndrome may not serve well in some of the most populous regions of the world. This is not a call for rejection of the stage-transition approach in future genetics research on drug dependence. However, it is a call for an examination of some fairly simple phenotypes that may prove to be measured more readily in the cross-national context and in diverse socio-cultural milieu. As illustrated above in Figure 1, these concepts are pertinent to stages of the drug dependence process that occur prior to the formation of the drug dependence syndrome (before dependence begins to drive the count processes), as well as to later stages. As discussed in Section 3 of the article, the earliest count processes involve (a) whether a drug is consumed for a second time, and (b) the persistence and rate ratio parameters of the ZIP regression model, all of which can be studied via fairly elementary survey items on month and day of first drug use, month and day of last drug use, and how many days of use have occurred in recent intervals.

The scale of the NSDUH, with thousands of newly incident recent onset drug users identified every 1-2 years, makes it ripe for rapid development of a future elaboration that becomes genetically informative in this fashion. In order not to disrupt the time trends, the elaboration can involve (a) leaving the NSDUH sampling, recruitment, and assessment modules alone, unchanged, and (b) offering the NSDUH participants an opportunity to participate in a subsequent assessment module and a research activity that fills in gaps not yet addressed in the NSDUH assessment modules. Our research group already has completed a pilot study that simulates how this subsequent assessment module and research activity might be completed, using tobacco smoking and tobacco dependence as the primary objects of study.

First, upon completion of the initial assessment modules for our pilot community survey, the participant was asked about willingness to participate in a sub-study that would involve: (a) answering an additional brief questionnaire survey on a form that is numbered but does not involve recording names, addresses, etc., (b) providing an anonymous saliva specimen in a bottle on which is recorded the same form number, and (c) using that same form number to log in to an internet survey site for completion of a brief survey on additional topics. That is, the form number was drawn at random by the participant and was recorded on the sub-study survey form and on the saliva specimen bottle, and was keyed in by the participant as part of the online survey (so as to allow linkage to the saliva specimen results). However, that form

number is known only to the participant, and remains unknown to all field research staff members.

In the pilot study, participants were allowed to complete all of these operations at the end of an initial community survey session within the sampled dwelling unit (typically a household). The field staff members carried wireless-enabled laptops that allowed the participant to access the internet survey website from the home. Some participants preferred to log in from a cybercafé, from school, the workplace, library, or other internet-connected devices, and did not take advantage of immediate access to the online survey website.

In the request to provide an anonymous saliva specimen, the participant in our pilot study was informed that a reinforcer of small value (US\$1.00 in our pilot study) would be delivered immediately, provided the participant was willing to donate the saliva for a bioassay to allow confirmation of a self-report of recent tobacco smoking, and that an additional reinforcer of small value (US\$2.00 in our pilot study) would be provided if the participant would consent to anonymous genotyping of the saliva. Participants who consented to donate the saliva specimen drew a plastic envelope from a bundle of bottle-containing envelopes, each one with a form number label and brief questionnaire about recent smoking and aspects of the tobacco dependence phenotypes (e.g., number of days smoked during the past 30 day interval). The consenting participants donated the specimen, marked the questionnaire responses on the envelope label, and also marked whether the bioassay, the genotyping, or both were to be permitted. (The participant can be permitted to take these materials to a nearby mailbox for mailing to the study office. In our pilot study, it was more typical for the participants to agree to drop the envelope into a compartment in a satchel carried by the field staff member.)

A reinforcer-linked encouragement design was used to gain partial experimental control over participation levels for the online survey. In specific, in the random draw of study form numbers, the participants concurrently were drawing at random from a gradient of reinforcers to be delivered upon completion of the online survey. These values, across a range from \$5 to \$50, were redeemable in form of retail gift cards. The research staff arranged to have the national retail chain store gift card distributed at the time of the face-to-face contact with the participant, pre-loaded with US\$5.00 as a token of thanks for initial steps in the survey process. The staff also set up an arrangement so that the appropriate reinforcer value was electronically deposited in the gift card account once the participant completed the online survey log-in. As reported elsewhere (Reed et al., under review; Rios-Bedoya et al., under review), the participant's random draw across a gradient of reinforcer values, coupled with the use of retail store gift cards and discount coupons can produce a fairly regular gradient of online survey participation level, as might be predicted from the behavior analytic perspective, with participation at the 75% level and sometimes as large as 90%-95% at the highest reinforcer values. The research approach involves making use of the randomly drawn reinforcer value as an instrument to help correct the observed distributions, based on the members of the sample who participate, so as to bring them into greater balance with the distributions that might have been observed if 100% participation had been secured. That is, the goal of this approach to a randomized reinforcer gradient in the encouragement design is not simply to find a fixed reinforcer value that can be used in future surveys to optimize the participation level. The goal is to harness the randomly drawn gradient of reinforcer values to improve the evidence and inferences from the study data.

In sum, the community pilot study approach our research group has designed is one that can be adapted in a sub-study for consenting participants after completion of the NSDUH, without damage to NSDUH participation levels. At the margin, the time required for the sub-study recruitment and completion at the end of the field research staff member's visit for

the NSDUH can be constrained to be short. All that is required is explanation of what is requested via an institutional review board-approved consent process, followed by the random draw and (a) completion of the brief survey form, (b) donation of the saliva specimen, and (c) if the participant wishes, mailing of the sub-study materials back to a research office. Research activities (a) and (b) require no more than a few minutes, and this interval depends heavily upon how much the IRB requires to be stated in a disclosure explanation as part of the consent process for an anonymous sub-study. As noted above, most participants do not choose the option of walking to the mailbox.

In proposing this type of future elaboration of the NSDUH, we are mindful that SAMHSA already has allowed other extramural investigators to conduct sub-studies after conclusion of the within-DU assessment, with solicitation of anonymous urine specimens for confirmation of self-reported recent drug use status. Hence, the precedent for this type of post-NSDUH sub-study with donation of a biological specimen for toxicological assays has already been established.

Moreover, we are mindful of the potential scientific significance of the annual repetition of the NSDUH sampling, recruitment, and standardized assessments, with samples producing as 60-70 thousand respondents each year, thousands of whom are self-identified as recent-onset drug users each year. In that genome wide association studies of drug dependence phenotypes generally require large samples, with provisions for internal replication samples, the NSDUH might be regarded as a GWAS resource for NIH-funded intramural or extramural research that now is squandered with each passing year this opportunity remains unexploited.

Our re-framing of the drug dependence phenotype in relation to relatively simple count processes in the early stages of the drug dependence process is one that is conducive to NSDUH pilot studies before this type of annual study operation is brought to scale. We originally started to think through the option of count process models in the earliest stages of drug involvement when we were facing the DSM-IV drug dependence measurement difficulties in some of the WMHS site-specific latent structure analyses. Nonetheless, a shift to a complementary set of easy-to-measure drug dependence phenotypes, as required for the count process research, also reduces the assessment burdens for the proposed elaboration of the NSDUH. Once there is convincing evidence that an appropriately calibrated reinforcer value can produce desirable online survey participation values at the level observed in our pilot studies with this encouragement design, it should be possible to add standardized assessments of the clinical features of drug dependence to the NSDUH sub-study online survey (in addition to the simpler measurements of the count process). Moreover, once experience with NSDUH sub-study recruitment on a larger scale has been gained, it may be possible to devise ways to conduct longitudinal follow up studies that go beyond a single post-NSDUH online survey assessment, that are extended to other genetically informative household or family members (e.g., the twin pairs or co-twins represented in each sampled dwelling unit; the parents and grandparents of the adolescent and young adult participants, as in 36), and that allow sub-sampling on the basis of the core NSDUH assessments (e.g., to permit a focus on a sample of recent-onset drug users and never user controls, based on what the participants have reported during the NSDUH ACASI assessment).

At present, the NSDUH study operation produces a remarkable national treasure in the form of a well-studied nationally representative sample of adolescent and adult community participants, characterized in relation to a host of phenotypic characteristics and other details that are pertinent in genetic epidemiology research and general genetics research on drug dependence. Our future research in this domain can be enriched by taking greater advantage of that national treasure along the lines sketched in this article.

Once NSDUH protocols have been elaborated in this direction, it should be possible to create a reiteration of the World Mental Health Surveys, adapting these same protocols for the international and cross-national context. Novel research work along these lines will produce the truly global perspective on genetic determinants and environmental modulation of gene expression. In time, the result may be practical applications and the betterment of global health in the domain of drug dependence and related phenotypes.

Acknowledgments

This work was supported by NIDA awards K05DA015799 and R01DA016558, as well as Michigan State University research funds from the Provost's Office). Thanks to Mirjana Radovanovic, Hui Cheng, and Brian Fairman for research assistance. The United States Substance Abuse and Mental Health Services Administration Office of Applied Studies, which administers the NSDUH and arranges for timely release of the NSDUH public use datasets, merit gratitude, as do Ronald Kessler of Harvard University and the other principal collaborators and staff members of the World Mental Health Surveys Consortium.

References

1. Gershon ES, Kelsoe JR, Kendler KS, Watson JD. Editorial: A Scientific Opportunity. *Science*, New Series. 2001; 294(5544):957.
2. Reboussin BA, Anthony JC. Latent class marginal regression models for modelling youthful drug involvement and its suspected influences. *Statistics in Medicine*. 2001; 20(4):623–639. [PubMed: 11223904]
3. Storr CL, Zhou H, Liang KY, Anthony JC. Empirically derived latent classes of tobacco dependence syndromes observed in recent-onset tobacco smokers: epidemiological evidence from a national probability sample survey. *Nicotine Tob Res*. 2004; 6(3):533–545. [PubMed: 15203787]
4. Storr CL, Reboussin BA, Anthony JC. The Fagerström test for nicotine dependence: a comparison of standard scoring and latent class analysis approaches. *Drug Alcohol Depend*. 2005; 80(2):241–250. [PubMed: 15908142]
5. Reboussin BA, Anthony JC. Is there epidemiological evidence to support the idea that a cocaine dependence syndrome emerges soon after onset of cocaine use? *Neuropsychopharmacology*. 2006; 31(9):2055–2064. [PubMed: 16482089]
6. Van Etten ML, Anthony JC. Comparative epidemiology of initial drug opportunities and transitions to first use: marijuana, cocaine, hallucinogens and heroin. *Drug Alcohol Depend*. 1999; 54(2):117–125. [PubMed: 10217551]
7. Delva J, Van Etten ML, González GB, et al. First opportunities to try drugs and the transition to first drug use: evidence from a national school survey in Panama. *Subst Use Misuse*. 1999; 34(10):1451–1467. [PubMed: 10446769]
8. Van Etten ML, Anthony JC. Male-female differences in transitions from first drug opportunity to first use: searching for subgroup variation by age, race, region, and urban status. *J Womens Health Gend Based Med*. 2001; 10(8):797–804. [PubMed: 11703892]
9. Caris L, Wagner FA, Ríos-Bedoya CF, Anthony JC. Opportunities to use drugs and stages of drug involvement outside the United States: Evidence from the Republic of Chile. *Drug Alcohol Depend*. 2009; 102(1-3):30–34. [PubMed: 19328635]
10. Wagner FA, Anthony JC. Male-female differences in the risk of progression from first use to dependence upon cannabis, cocaine, and alcohol. *Drug Alcohol Depend*. 2007; 86(2-3):191–198. [PubMed: 17029825]
11. Neumark YD, Van Etten ML, Anthony JC. “Drug dependence” and death: survival analysis of the Baltimore ECA sample from 1981 to 1995. *Subst Use Misuse*. 2000; 35(3):313–327. [PubMed: 10714449]
12. Helzer JE, Robins LN, Davis DH. Antecedents of narcotic use and addiction. A study of 898 Vietnam veterans. *Drug Alcohol Depend*. 1976; 1(3):183–190. [PubMed: 1017369]
13. Brady JV. Issues in human drug abuse liability testing: Overview and prospects for the future. 1989:357–370. In:

14. Koob GF, LeMoal M. Drug abuse: hedonic homeostatic dysregulation. *Science*. 1997; 278(5335): 52–58. [PubMed: 9311926]
15. Johnson C, Drgon T, Liu QR, et al. Genome wide association for substance dependence: convergent results from epidemiologic and research volunteer samples. *BMC Med Genet*. 2008; 9:113–123. [PubMed: 19094236]
16. National Cancer Institute. Bethesda, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute; 2009 [17 November 2009]. Phenotypes and Endophenotypes: Foundations for Genetic Studies of Nicotine Use and Dependence Tobacco Control Monograph No. 20. NIH Publication No. 09-6366, August 2009; http://cancercontrol.cancer.gov/tcrb/monographs/20/m20_entire.pdf
17. Churchill FB. William Johannsen and the Genotype Concept. *Journal of the History of Biology*. 1974; 7:5–30. [PubMed: 11610096]
18. Roll-Hansen N. Sources of Wilhelm Johannsen's Genotype Theory. *Journal of the History of Biology*. 2009; 42(3):457–493. [PubMed: 20027784]
19. Demyttenaere K, Bruffaerts R, Posada-Villa J, et al. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA*. 2004; 2:2581–2590. [PubMed: 15173149]
20. Degenhardt L, Bohnert KM, Anthony JC. Assessment of cocaine and other drug dependence in the general population: “gated” versus “ungated” approaches. *Drug Alcohol Depend*. 2008; 93(3): 227–232. [PubMed: 18065161]
21. Kessler RC, Aguilar-Gaxiola S, Alonso J, et al. The global burden of mental disorders: An update from the WHO World Mental Health (WMH) surveys. *Epidemiol Psychiatr Soc*. 2009; 18(1):23–33. [PubMed: 19378696]
22. Kessler RC, Üstün TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods Psychiatr Res*. 2004; 13(2):93–121. [PubMed: 15297906]
23. Degenhardt L, Bohnert KM, Anthony JC. Case ascertainment of alcohol dependence in general population surveys: ‘gated’ versus ‘ungated’ approaches. *Int J Methods Psychiatr Res*. 2007a; 16(3):111–123. [PubMed: 17703472]
24. Degenhardt L, Cheng H, Anthony JC. Assessing cannabis dependence in community surveys: methodological issues. *Int J Methods Psychiatr Res*. 2007b; 16(2):43–51. [PubMed: 17623384]
25. Üstün B, Compton W, Mager D, et al. WHO Study on the reliability and validity of the alcohol and drug use disorder instruments: overview of methods and results. *Drug Alcohol Depend*. 1997; 47(3):161–169. [PubMed: 9306042]
26. Muthén, LK.; Muthén, BO. *Mplus User's Guide*. Fifth. Los Angeles, CA: Muthén & Muthén; 1998-2007.
27. Anthony JC, Ritter CJ, Von Korff MR, et al. Descriptive epidemiology of adult cocaine use in four U.S. communities. *NIDA Res Monogr*. 1986; 67:283–289. [PubMed: 3092079]
28. Anthony JC, Chen CY, Storr CL. Drug dependence epidemiology. *Clinical Neuroscience Research*. 2005; 5(2-4):55–68.
29. Anthony JC, Petronis KR. Early-onset drug use and risk of later drug problems. *Drug Alcohol Depend*. 1995; 40(1):9–15. [PubMed: 8746919]
30. Breslau N, Peterson EL. Smoking cessation in young adults: age at initiation of cigarette smoking and other suspected influences. *Am J Public Health*. 1996; 86(2):214–220. [PubMed: 8633738]
31. Bierut LJ, Madden PA, Breslau N, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet*. 2007; 16(1):24–35. [PubMed: 17158188]
32. Anthony JC, Warner LA, Kessler RC. Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: Basic findings from the National Comorbidity Survey. *Experimental and Clinical Psychopharmacology*. 1994; 2:244–268.
33. Hilbe, JM. *Negative Binomial Regression*. New York: Cambridge University Press; 2007. p. 251
34. Long, JS.; Freese, J. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd. College Station, TX: Stata Press; 2006. p. 527

35. Chen YC, Peng GS, Tsao TP, et al. Pharmacokinetic and pharmacodynamic basis for overcoming acetaldehyde-induced adverse reaction in Asian alcoholics, heterozygous for the variant ALDH2*2 gene allele. *Pharmacogenet Genomics*. 2009; 19(8):588–599. [PubMed: 19584771]
36. Kendler KS, Myers J, Potter J, Opalesky J. A Web-Based Study of Personality, Psychopathology and Substance Use in Twin, Other Relative and Relationship Pairs. *Twin Research and Human Genetics*. 2009; 12(2):137–141. [PubMed: 19335183]

1. ONSET OF DRUG USE
2. COUNT OF DRUG EXPERIENCES
3. CLINICAL FEATURES FORM
4. SYNDROME COALESCENCE
5. SECONDARY COMPLICATIONS

Figure 1.

Violation of the acyclic assumption. Often, there is an insidious onset of the drug dependence syndrome, once use starts. Plus, as the dependence develops, feedback loops appear, such that the choice to quit becomes more difficult. The count of drug experiences drives the dependence process, but in addition, the count is itself influenced by the dependence process. (Copyright, J.C. Anthony, 2009. Permission granted for use of this figure in J.C. Anthony *Addiction Reviews* article within which it is contained.)

IRT Discrimination Parameter Estimates

Figure 2.

Estimated Item Response Theory Discrimination Parameter Estimates from four sites of the World Mental Health Survey Consortium, based on standardized field survey assessments of the DSM-IV drug dependence syndrome with respect to internationally regulated drugs (primarily cannabis). Data from large multi-stage nationally representative population samples in each of six de-identified sites, 2002-2007.

Figure 3.

The estimated risk of developing clinical features associated with tobacco dependence among recent-onset tobacco smokers, by age of onset (11-17 years versus >17 years), within 24 months after first onset of smoking. Data are from the National Household Survey on Drug Abuse and National Surveys on Drug Use and Health, 2000-2002 (Citation 28). The clinical feature abbreviations are explained in the text. [This figure was adapted and re-drawn from a previously published figure for which this article's author is responsible. Anthony et al., 2005].

Table 1

Pattern of Subgroup Variation in the Persistence and Rate Ratio Parameter Estimates from Zero-Inflated Poisson Regression Models Fit to Days of Drinking in the 30 Days Prior To Survey Assessments Completed Within 24 Months After Onset of Drinking. Data from National Surveys on Drug Use and Health, United States nationally representative community samples of non-institutionalized citizens age 12 years and older, 2002-2007.

Psychoactive Drug Compounds Under Study -- Recent Onset Users of:	Estimated Persistence Parameter	Estimated Rate Ratio Parameter, Conditional on the Persistence Outcome
Alcoholic beverages	Asian-Americans (AA) more likely to be in 'Subsequently Always a Never Drinker' ($p < 0.05$)	No difference between Asian-Americans and non-Hispanic Whites (nHW)
Cannabis	AA = nHW	AA = nHW
Cocaine	AA = nHW	AA = nHW
Tobacco	AA = nHW	AA = nHW

Data from Meyer et al., under review; Barondess et al., under review; Mainampally et al., under review; Anthony et al., under review.