Behavioral Responses to Surveys

About Nicotine Dependence

by

Glenn W. Harrison

September 2017

C.V. Starr Chair of Risk Management & Insurance, and
Director, Center for the Economic Analysis of Risk,
Department of Risk Management & Insurance,
Robinson College of Business, Georgia State University, USA

Mailing address: P.O. Box 4036, Atlanta, GA 30302-4036, USA

Forthcoming, *Health Economics*

Declaration of Competing Interests: Harrison has served as a testifying expert witness for plaintiffs in litigation against tobacco companies in the United States and Canada for 20 years, and continues to do so. His testimony is on health care costs to governments associated with tobacco-related diseases. He has received compensation for the research and time supporting this testimony.

Abstract: Behavioral responses to surveys can significantly affect inferences about population prevalence unless correctly modeled statistically. An important case study is the prevalence of nicotine dependence, a formal psychiatric disorder satisfying clinical criteria. Data from the National Epidemiologic Survey on Alcohol and Related Conditions in the United States are used, along with a flexible semi-nonparametric sample selection model. Corrections for sample selection responses to "gateway" survey questions lead to significantly higher estimates of the prevalence of nicotine dependence among current daily smokers. These corrections also imply even higher levels of the decades-long and lifetime-long persistence of nicotine dependence after the onset of smoking.

Key Words: nicotine dependence, sample selection bias, population prevalence, survey methods.

1. Introduction

The population prevalence of Nicotine Dependence is measured using surveys that ask questions about the criteria defining a diagnosable psychiatric disorder. These survey questions are screens designed to detect individuals who might clinically "present" and then meet criteria for diagnosis of the psychiatric disorder. These survey screens generally use questions defined directly or approximately in terms of *DSM-IV* and *DSM-5* (American Psychiatric Association [1994][2013]) clinical criteria for Nicotine Dependence (ND).[1]

A significant statistical and public policy issue arises from the difficulty of drawing inferences about population prevalence when one attempts to account for possible *behavioral responses* to "trigger," "gateway" or "diagnostic stem" questions. These are one, two or three questions which ask if the individual has had some general experience with the disorder. For instance, and using the survey evaluated below: for tobacco, and only referencing cigarettes, "In your entire life, have you ever smoked at least 100 cigarettes?" Anyone that responded "no" to the trigger question was not asked any of the questions about the criteria for ND, and is automatically classified as not being ND, in some applications, and as having no detectable risk of being ND in other applications. However these survey respondents are classified, they do not contribute in any form to the estimate of the "at risk" population prevalence of ND.

What is the impact of these behavioral responses to surveys on the inferences drawn about dependence on nicotine? These questions can be answered, to some extent, by the use of corrections for "sample selection bias." Our primary objective is to demonstrate how to apply these corrections, and to assess their significance for inferences about population prevalence and the persistence of nicotine dependence over time.

---

[1] Nicotine Dependence has *DSM-IV* code 305.10. This is the same as the ICD9-CM code, which is true for most psychiatric disorders. In *DSM-5* the disorder is renamed Tobacco Use Disorder.

It is clear why these trigger questions are used. They save on valuable respondent time in a potentially long survey instrument. They also avoid asking questions that would not make sense, or would have to be phrased speculatively in the subjunctive mood.

However, it is equally clear, on reflection, why behavioral responses to these trigger questions should not be taken at face value. The *potential* for sample selection bias arises when there is some systematic factor explaining why someone might not want to participate in the full set of questions, and therefore deliberately or subconsciously selects out of that full set by answering a certain way in response to the trigger question.[2] Sometimes this potential leads to no difference in inferences from the observed sample: for instance, if respondents want to spend more time in a face-to-face interview with more attractive interviewers, and the attractiveness level of interviewers is random, there will be no *a priori* reason to expect an effect on inferences about addiction risks. On the other hand, if someone wants to hide their ND, for example, they might reasonably choose to lie in response to the trigger question.[3] Certain aspects of nicotine addiction might understandably be something that the respondent wanted to hide, was just disinclined to talk about, or was in denial of.

Furthermore, many patients with diagnosable medical and psychiatric conditions simply do not present for clinical evaluation. One proxy for that decision not to clinically present may, plausibly, be an unwillingness to open the topic for discussion, reflected in a "no" response to the

---

[2] Hernán, Hernández-Diaz and Robins [2004] survey the many types of selection bias considered in epidemiology, and provide a general causal framework. The selection bias of concern here is a mixture of what they call "nonresponse bias/missing data bias," "volunteer bias/self-selection bias," and "health worker bias" (p. 618). Various statistical correction methods are discussed in major epidemiology texts, such as Rothman, Greenland and Lash [2012; ch. 19]. There appear to be no applications of epidemiological corrections for these biases to general population surveys with trigger questions.

[3] The problem becomes even more severe for certain disorders. For example, for the disorder Pathological Gambling, 312.31, the trigger question is, "Have you ever gambled at least 5 times in any one year of your life?" However, hiding gambling problems is explicit in one of the criteria used in the full set of questions for determining the extent to which someone is at risk for a gambling disorder. Similarly, the disorder Social Phobia, 300.23 has two criteria questions that ask if the respondent has ever had a strong fear or avoidance of "being interviewed" or "having conversations with people you don't know well." The survey considered here used face-to-face interviews.

trigger question. The social acceptance of smoking behavior has varied over the years, and in different cultures, and might be expected to vary within a sampled population.[4]

2. Methods

A. Data

There exist many survey instruments designed to measure ND, such as the Fagerström Test for Nicotine Dependence (e.g., Heatherton et al. [1991]). These survey instruments often reflect one or more of the criteria needed to define ND according to the *DSM-IV*, and may be useful screening devices for some cessation purposes, but we focus attention on the formal, complete *DSM-IV* criteria when referring to ND. The *only* major source of data to measure ND at a population level appears to be the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) in the United States. The first wave of NESARC was conducted in 2000/1, and had a sample of 43,093 individuals. This survey has generated many estimates of the prevalence of substance use disorders, such as Grant et al. [2004a][2004b].

Current ND in the past 12 months is defined in the *DSM-IV* by the presence of 3 or more of 7 criteria within the past 12 months, for *someone that is currently a daily smoker*. This last qualification is important, and natural: someone cannot begin to be classified as currently nicotine dependent unless they are actually smoking daily. Using survey instruments developed by Grant et al. [2003], the NESARC contained a number of questions that mapped directly into these criteria.

B. Statistical Procedure

The sample selection models developed by Heckman [1976][1979] provide one statistical

[4] There are a several ways to design surveys to mitigate sample selection biases, discussed in an Appendix.

method for evaluating the potential importance of trigger questions. They require the researcher to specify a sample selection process, characterizing which respondents appear in the main survey module for a disorder and which do not. Typically this is a binary matter, so one can specify this process of responding to the trigger question with a probit model. In our case this consists of a binary choice statistical model explaining whether someone responded affirmatively to the trigger question and was then asked the questions about the criteria for the disorder.

In the original setting studied by Heckman [1976][1979] the main data generating process of interest, and potentially subject to sample selection bias, had a dependent variable that was continuous, and the statistical specification was Ordinary Least Squares. However, a more natural specification for ND is to think of the classification in terms of several ordered categories, each indicating that the individual has a greater risk of being ND. This specification allows one to remain agnostic about the conjectural "bright line" between addiction and non-addiction that a binary classification requires, and is consistent with increased diagnostic emphases on "severity" in *DSM-5*.

For ND, respondents can be classified into 4 categories: **No Risk** individuals report no *DSM-IV* criteria or not being asked about them; **At Risk** individuals report 1 or 2 *DSM-IV* criteria; **High Risk** individuals report 3 or 4 *DSM-IV* criteria; and **Severe Risk** individuals report 5 or more *DSM-IV* criteria.[5] The last two categories define ND in *DSM-IV*.

Previous statistical evaluations of hierarchies of this kind have not formally recognized the *ordered* nature of the categories used in standard survey screens, which are derived directly from

[5] Most of the DSM criteria include the requirement that the symptoms be "clinically significant." This is normally identified by questions asking if the symptom(s) led to any contacts with medical professionals, use of medication more than once, or led to interference with "life or activities." For reasons of survey efficiency, these questions are normally asked only if the respondent meets some threshold level of symptoms. Hence one must be careful to recognize that anyone that has met fewer than the threshold level of symptoms will not have been asked about clinical significance (and, more generally, that these thresholds can be applied differently across general surveys, leading to apparent discrepancies in prevalence estimates, as stressed by Narrow, Rae, Robins and Reiger [2002]). This exclusion criteria is also only asked in surveys if someone met the threshold level of symptoms.

clinical screens. When several categories are ordered there are appropriate estimation procedures that use this information. The most popular are ordered probit models in which a latent index is estimated with "cut points" to identify the categories. We employ a flexible Semi-Nonparametric (SNP) version of this type of ordered response model, developed by Stewart [2004] and extended by De Luca and Perotti [2011] to allow for sample selection corrections.

One important assumption in the standard sample selection model is to specify some structure for the errors of the two equations, the sample selection equation and the main survey question. If both equations are modeled with probit specifications, for example, a natural assumption is that the errors are bivariate normal. We assume instead the SNP approach due to Gallant and Nychka [1987], which approximates the bivariate density function of the errors by a Hermite polynomial expansion.[6]

Following De Luca and Perotti [2011; p. 215], the ordered probit sample selection model can be defined in three equations:

$$Y_j^* = \beta_j^\top X_j + U_j \qquad j = 1, 2 \qquad (1)$$

$$Y_1 = \mathbf{I}(Y_1^* \geq 0) \qquad (2)$$

$$Y_2 = \sum_{h=0\ldots H} h\, \mathbf{I}(\alpha_h < Y_2^* \leq \alpha_{h+1}) \qquad \text{if } Y_1 = 1 \qquad (3)$$

where Y_1^* is a continuous latent variable for the sample selection equation, Y_2^* is a continuous latent variable for the risk of nicotine dependence, β_j denotes k_j vectors of parameters to be estimated, X_j denotes k_j vectors of exogenous variables, the U_j are random errors, $\mathbf{I}(\cdot)$ is the indicator function, Y_1 is the binary variable indicating the observed sample when $Y_1 = 1$, Y_2 is the observed level of

[6] This SNP approach is computationally less intensive than comparable approaches based on the estimation of kernel densities. There is some evidence from Stewart [2005] and De Luca [2008] that this SNP approach has good finite sample performance when compared to conventional parametric alternatives and other SNP estimators. Stewart [2004; §3] provides an excellent discussion of the mild regularity conditions required for the SNP approximation to be valid, and the manner in which it is implemented so as to ensure that a special case is the parametric (ordered) probit specification. An appendix (online) documents the model specification more formally, and identification restrictions.

nicotine dependence, H+1 denotes the ordered categories of nicotine dependence, and (α_0, ..., α_h, α_{h+1}, ..., α_H) are thresholds to be estimated, with $\alpha_0 = -\infty$, $\alpha_h < \alpha_{h+1}$ and $\alpha_H = \infty$.

Equation (2) defines the sample selection process by which we observe the sample for which $Y_1 = 1$, and by itself is just a probit equation. Equation (3) defines the ordered probit, conditional on sample selection, which means conditional on responding affirmatively to the trigger question for nicotine dependence. The H cutpoints (α_0, ..., α_h, α_{h+1}, ..., α_H), to be estimated, define H+1 intervals over the latent variable Y_2^*. The correlation of the latent regressions errors U_1 and U_2 determines selectivity effects. If this correlation is positive (negative) then it means that unobservables have the same (opposite) effect on selection and the risk of nicotine dependence.[7]

An important assumption in the sample selection model, said to be "good for identification," is to find variables that explain sample selection but that *a priori* do not explain the main outcome. In many expositions one sees the comment that in the absence of these "exclusion restrictions" the sample selection model is "problematic." Often this is a major empirical challenge, since it can be hard to exclude something from potentially affecting the main variable of interest, but to include it as likely to affect sample selection. In epidemiology, for instance, a spirited defence[8] of the use of sample selection corrections to estimates of HIV prevalence in Bärnighausen et al. [2011a] came from Bärnighausen et al. [2011b] on the grounds that they had access to ideal exclusionary

[7] The traditional parametric specific of the model assumes that the errors U_1 and U_2 follow a bivariate Normal distribution with zero means, unit variances, and a correlation coefficient ϱ. The SNP innovation is to approximate the marginal distribution functions of U_1 and U_2, and their joint distribution function. The approximation starts with an approximation of the joint density by the product of a standardized normal density for U_1; a standardized normal density for U_2; a polynomial of order R in U_1 and U_2, with R×R polynomial coefficients to be estimated; and a normalization factor. Once this joint density is approximated, one can use it to approximate the marginal distribution functions of U_1 and U_2 (De Luca [2008]). The fact that the standardized normal densities are used for the first two terms of this approximation means that a special case of the SNP specification is the parametric specification, allowing a direct test of the hypothesis that the SNP estimates are the same as the parametric estimates.

[8] Criticisms were raised by Geneletti, Mason and Best [2011] in response to epidemiological applications of corrections for sample selection by Chaix et al. [2011] and Bärnighausen et al. [2011a].

restrictions: the identity of the survey interviewer. We agree that this exclusion restriction is an attractive and reasonably general one, but it is not universally applicable.

What is *particularly* "problematic" in the absence of *a priori* convincing exclusion restrictions is that one must rely on having the right econometric specification of the error distribution if the sample selection model is to correct for sample selection bias. This specification refers to the parametric nature of the assumed bivariate normality of errors. But the importance of having the right specification of the error distribution also applies even when one does have exclusion restrictions. As it happens, a formal identification requirement of the SNP specification is that one have some exogenous variables in the sample selection equation that are not in the main equation for nicotine dependence. In this respect, the SNP specification is more restrictive than the parametric specification, although the restriction is easy to meet in the case of trigger questions being the source of sample selectivity.

As it happens, there are ways to construct exclusion restrictions in NESARC that have some *a priori* credibility. For instance, we know the day of the week that the interview was conducted on, and can condition on Friday, Saturday or Sunday interviews as generating differential response. We also know how many trigger questions for other disorders a subject had answered affirmatively by the time the addiction trigger questions were asked, as one measure of how much time had been taken by that stage of the interview. Additional characteristics of the individual are available from baseline questions, and can be used to identify the sample selection equation.

Core covariates in both equations of the sample selection model include binary variables for gender, ethnicity (black, hispanic), age (aged between 18 and 29, aged between 45 and 64, aged 65 and over), marital separation (due to being widowed, divorced or separated), region (Midwest, South, West), education (completed high school or GED, completed some college, completed a college degree, completed a graduate degree), personal income (below $20,000; between $20,000 and

$35,000; $70,000 or more), and whether any welfare payments had been received. Additional covariates for the sample selection equation include height, weight, day of the interview (Friday, Saturday, Sunday), and 21 questions about recent events in the life of the respondent.[9] An online appendix describes all covariates used, along with summary descriptive statistics.

3. Results

Figure 1 reports estimates of population prevalence for ND that ignore sample selection and estimates that correct for it.[10] The fractions of the population from the raw data found in each *DSM-IV* response category are accurately recovered by the estimated ordered response model when we do not correct for sample selection. Hence we know that the base statistical model is not biased relative to the raw data, at least as we have binned it into ordered categories.

The general finding is that sample selection corrections increase the estimated population prevalence of individuals at risk of ND. As defined by the *DSM-IV*, ND prevalence increases from 12.7% to 35.8%. This result is not simply because the sample selection model predicts that more people will get through the gateway of the trigger question than the raw data implies, although it does predict that. The observed fraction being selected by their responses to the trigger question is 44%, and the predicted fraction from the sample selection model who would have been selected if they answered the trigger question accurately, according to the empirical specification, is 64%.[11] The

[9] These questions reflect how much time in the month prior to the interview the respondent had problems with work or regular daily activities due to physical health, emotional problems, or pain; how much time in the prior month the respondent felt calm, had a lot of energy, felt downhearted, or had physical or emotional problems interfere with social activities; and if, in the prior year, the respondent experienced the death of a family member or close friend, experienced a family member or close friend have a serious injury, moved or had someone new live with them, was fired or laid off, was unemployed for more than a month, had work trouble, changed jobs, broke off a steady relationship, had serious problems with a neighbor or relative, experienced a major financial crisis, had trouble with police, or was a victim of a crime.

[10] All computer code used to generate these estimates is available at http://cear.gsu.edu/gwh/.

[11] Because the predicted fraction to be selected exceeds the observed fraction, one might just assume that the selection equation is mis-specified, and this is the simple explanation for our findings of a higher

issue is also a matter of which *profile* of subjects is predicted to be selected. The sample selection model predicts *more* of the types of people predicted to flag *more DSM* criteria, and *fewer* of the type of people predicted to flag *fewer DSM* criteria. Thus sample selection is, as emphasized by Heckman [1976][1979], fundamentally an issue about allowing for unobserved heterogeneity.

Figures 2 displays the distribution of predictions, with and without sample selection corrections, as well as indicators of the statistical significance of the effect of sample selection. Consider the top left panel in Figure 2, for the "No Risk" category of ND. The Uncorrected distribution of predictions reflects the results of simulating 100 random draws for each NESARC respondent from the predicted marginal probability of No Risk, using the estimated SNP ordered probit model. Each random draw is from a normal distribution whose mean is the point estimate of the marginal probability for that subject, and whose standard deviation is the standard error of that point estimate, again for that subject. Thus the 100 random draws for each subject reflect individual-specific predictions, taking into account the statistical uncertainty of the prediction. The Corrected distribution of predictions is generated similarly, using the estimated SNP ordered probit model allowing for sample selection. Since there are 43,093 respondents to NESARC, each of the kernel densities in Figure 2 reflect 4,309,300 predictions.

These densities in Figure 2 allow one to see the average effects shown in Figure 1, the decrease in predicted No Risk respondents from 0.77 to 0.53, but also to visualize the precision of this difference. A *t*-test for each NESARC respondent generates a *p*-value for the hypothesis that the

---

prevalence of individuals at risk. However, the predicted probability of being selected in the sample selection model is the predicted sample conditional on covariates *plus an error term* for that selection equation. In the usual parametric sample selection specification this error term is *assumed* to be zero, so these observed and predicted fractions should be more or less the same. However, the semi-nonparametric specification does not assume this error term to be zero, as emphasized by DeLuca and Perotti [2011; p.218]. Hence the predicted fraction could be larger or smaller than the observed fraction. This point further illustrates how the sample selection model benefits from not having to impose a parametric stochastic structure.

predicted marginal probability is the same with and without sample selection corrections. The 90th, 95th and 99th percentiles of this distribution of 43,093 *p*-values are tabulated in the top-left panel of Figure 2. We find that the predicted *decrease* in No Risk is quantitatively large and statistically significant, in the sense that the 99th percentile of these *p*-values is 0.001 or lower.[12] There are statistically significant increases in the High Risk and Severe Risk categories. For High Risk the 95th percentile of *p*-values is 0.002 or lower, and for Severe Risk the 99th percentile of *p*-values is also 0.002 or lower. There is no change in the average prevalence of 0.10 for the At Risk category.

What implications do these corrections to the prevalence of ND have for our understanding of the etiology and correlates of ND? Two particularly important characteristics are the longevity of ND, and age of ND.

Figure 3 displays raw data from NESARC showing the probability of someone being assessed to be ND in the year of the survey in relation to the number of years since the person started daily smoking.[13] It is natural to start by looking at the onset of daily smoking, since a diagnosis of ND require than the individual be a current, daily smoker in the year of the survey. These data show that in the first 10 years after starting smoking more than 40% of current daily smokers are ND. This fraction declines slowly and steadily, but remains roughly at 20% even after 40 years since the onset of smoking. Only after 75 years since the onset of smoking does the probability of being a current smoker and being ND drop to zero. An obvious factor leading to this eventual decline is the significantly greater mortality risk that a daily smoker faces.

Figure 3 also displays the predicted probability of ND, correcting for sample selection, in

---

[12] The percentile value is purely descriptive, as a summary statistic for 43,093 *p*-values. The *p*-value is the inferential statistic.

[13] This is the "jagged" thin line in Figure 3. Even with a sample size as large as the NESARC, there are relatively small samples for each year since the onset of smoking, resulting in some sampling variability around a clear trend.

relation to the number of years since the person started daily smoking. As expected from the general increase in the prevalence of ND, this probability is uniformly higher. The lower bound of the 95% confidence interval on this prediction includes the raw data for the first 10 to 15 years since the onset of smoking, but is otherwise above the raw data. These data, both the raw data and the corrected predictions, provide a stark reminder of the persistence of ND, measured in decades and (tragically shortened) lifetimes.

Figure 4 similarly shows the raw data on ND and predicted probability of ND in relation to the age of the current daily smoker. Remarkably, even at the age of 20 we see high rates of ND around 50% to 60%. One reason for high ND at such an early age is that current daily smokers started smoking much earlier in life on average: in the NESARC, at 15.7 in terms of any smoking, and at 18.3 for daily smoking. The longevity of ND with respect to age in Figure 4 follows from the longevity of ND with respect to the number of years smoking, but again demonstrates the long hold that ND has on those exposed to cigarettes.

4. Discussion

How might one mitigate some of the effects on prevalence estimates of survey screens for (psychiatric) disorders that use trigger questions, whatever the form of the question?

First, if possible one could design surveys that do not naively assume that trigger questions lead to no sample selection bias, by asking all DSM-related questions irrespective of the trigger response. This may mean that surveys have to be narrower in scope than the NESARC, given the time constraints implied. Or one could ask all DSM-related questions for one or two disorders selected at random. In each case some minor language changes would be needed, and one might sensibly ask the trigger questions at the end of the block, as in Harrison, Jessen, Lau and Ross [2017] for gambling disorders.

Second, where there is a need for some sort of trigger questions to avoid taking too much time in surveys, one can build in random treatments to make it easier to identify sample selection bias. These treatments might be conditions that affect the likelihood of someone participating in a full survey. An example is the use of financial incentives for participating in surveys, of the kind employed in some surveys and experiments (e.g., Harrison, Lau and Rutström [2009], Harrison, Lau and Yoo [2016]).[14]

These treatments might also be the acquisition of additional information that allows one to better identify the response bias that generates sample selectivity. An obvious example in the case of nicotine dependence is the vast literature comparing self-reports of smoking with objective measures that are known to be correlated with exposure to smoke (see Gorber et al. [2009] for a survey). An excellent example comes from the Canadian Health Measures Survey (CHMS). In this instance, "ever smokers" are asked quite detailed questions about their current and recent smoking behavior, and serum cotinine measurements made. These measurements are made between one day and 6 months after the initial survey responses. Using data from the 2007-2009 wave of the CHMS, Wong et al. [2012] show consistency between self-reports and objective measures, using a cutoff in cotinine of 50ng/ml to define "current smokers," which includes current daily smokers as well as current occasional smokers. So defined, the objective test has a sensitivity of 91.6% and a specificity of 98.3%, and there is no statistically significant difference between the two.[15] Information of this kind

[14] Another option would be to randomize the order in which disorders arise in the questionnaire. One of the significant determinants of sample selection, particularly for gambling, which appears near the end, is the number of trigger questions answered affirmatively to that point in the NESARC survey. Sample selection biases can arise from simple boredom or impatience, particularly with co-morbidities between psychiatric disorders.

[15] Sensitivity measures the percent of self-reported smokers who met the objective threshold for being declared a smoker, and sensitivity measures the percent of self-reported non-smokers who did not met the objective threshold for being declared a smoker.

could be used to generate statistical calibrating functions to better infer smoking prevalence[16] and to identify which characteristics are associated with reporting bias.[17]

Third, one could recruit subjects from an Administrative Registry, so that one can again better control for sample selection biases by knowing characteristics of all of those recruited. This is not a general option, since few non-Scandinavian countries have general registries, although recruiting from a Census may suffice.

5. Conclusions

Using data from the United States, corrections for behavioral responses to the use of survey trigger questions leads to significantly higher estimates of the prevalence of nicotine dependence among current daily smokers. Measurement of the population prevalence of addictions plays a critical role in public health policy and regulatory policy, particularly for substance dependence on products that are legally obtained without prescription. It should make a substantive difference to policy assessment and forecasts of consequences whether the fraction at risk of being dependent on nicotine is 12.7% or 36.2%. It is likely that these corrections would also affect estimates of comorbidities. Estimates of prevalence for addictions can also play a role in public policy towards the regulation of products of certain behaviors believed to be addictive. The long process of litigation against the tobacco industry in the United States (Derthick [2011]) has generated financial settlements of roughly 40% of attributable historical Medicaid treatment costs (Coller, Harrison and

[16] Gorber et al. [2008] demonstrate this possibility for self-reports of obesity, using data from the CHMS.

[17] The National Health and Nutrition Examination Survey (NHANES) for the United States has been used by Cawley and Choi [2015] in just this way. The NHANES conducts surveys and also collects samples from a subset of respondents, akin to the CHMS, and has been doing so every year since 1999. Using the self-report of any use in the last 5 days, and two objective thresholds, Cawley and Choi [2015] show that self-reports become more accurate as someone has greater formal education. The reason for this effect is not so relevant for our purposes, just that there is an effect that varies with a demographic characteristic of the individual. It is exactly this sort of *heterogeneity* in response bias that is driving selectivity effects.

McInnes [2002]).

Estimates of the persistence of ND in relation to age of onset of smoking play a role in evaluating the costs of smoking attributable expenditures. Starting in July 1997, Mississippi, Florida, Texas and Minnesota settled litigation with tobacco companies. In November 1998 the Attorneys General of 46 states in the United States entered into a Master Settlement Agreement (MSA) with the 4 largest manufacturers of cigarettes in the United States. The MSA (§IV) led to the public release of documents uncovered during the discovery process of the litigation, resulting in over 14 million documents being collated on the web, at www.industrydocumentslibrary.ucsf.edu/tobacco/, and subsequent academic research evaluating them. For the sake of argument, assume that as of 1998 public authorities in the United States had enough information to realize the extent of the misrepresentation of the risks of smoking that had occurred. Of course, it does not follow that this information instantly became "common knowledge" with the public, or that knowledge *per se* ends addiction.

Assume further that legal liability by tobacco companies for smoking attributable expenditures in the United States ended in 1998: for the litigating states in the United States, at least, this was a key provision (§XII) of the MSA. The evidence in Figure 3 suggests that some fraction of smoking after that date will be associated with (current daily) smokers who exhibit ND, and who started smoking before 1998.[18] It becomes a public policy and legal matter to determine if ND links the conduct and misrepresentation *prior* to 1998 to smoking attributable expenditures from *smoking that occurs after* 1998. The levels of the probability of ND in Figure 3, and the decades-long and lifetime-long persistence of ND after the onset of smoking, show that these are not likely to be trivial amounts. It would appear that the states that were parties to the MSA have no legal recourse

[18] This is apart from those who smoked prior to that date, and quit prior to that date, but who continue to experience additional health care expenditures associated with their (former) smoking.

to recover these smoking attributable expenditures, but of course they should factor into ongoing litigation in other countries.

Are these alternative prevalence estimates correct? It is self-evident that they do not come directly from data, and entail statistical inference resting on various assumptions. These sample selection models should be viewed as statistical "canaries in the cave," in the sense of pointing to potentially disastrous inferential conditions that warrant immediate investigation. In other words, and to put the inferential shoe on the other foot, if estimates of prevalence show great sensitivity to reasonable sample selection corrections with these assumptions, then one should not ignore that evidence because of the need for some statistical structure. Estimates derived directly from surveys of the kind considered here, which characterize virtually all population surveys of mental health disorders, simply cannot be taken at face value.

Acknowledgments

This analysis was prepared using a limited access data set obtained from the National Institute on Alcohol Abuse and Alcoholism (NIAA) and does not reflect the opinions or views of NIAAA or the U.S. Government. The use of these data has also been approved by the Institutional Review Board of Georgia State University. Valuable comments were received from three referees.

Figure 1: Predicted Prevalence of Nicotine Dependence for Current Daily Smokers, With and Without Sample Selection Correction

Estimated Probabilities using Semi-Nonparametric Ordered Response Model
Source: *National Epidemiological Survey on Alcohol & Related Conditions (NESARC)* of 2000/1
DSM-IV Nicotine Dependence changes from 12.7% to 35.8%

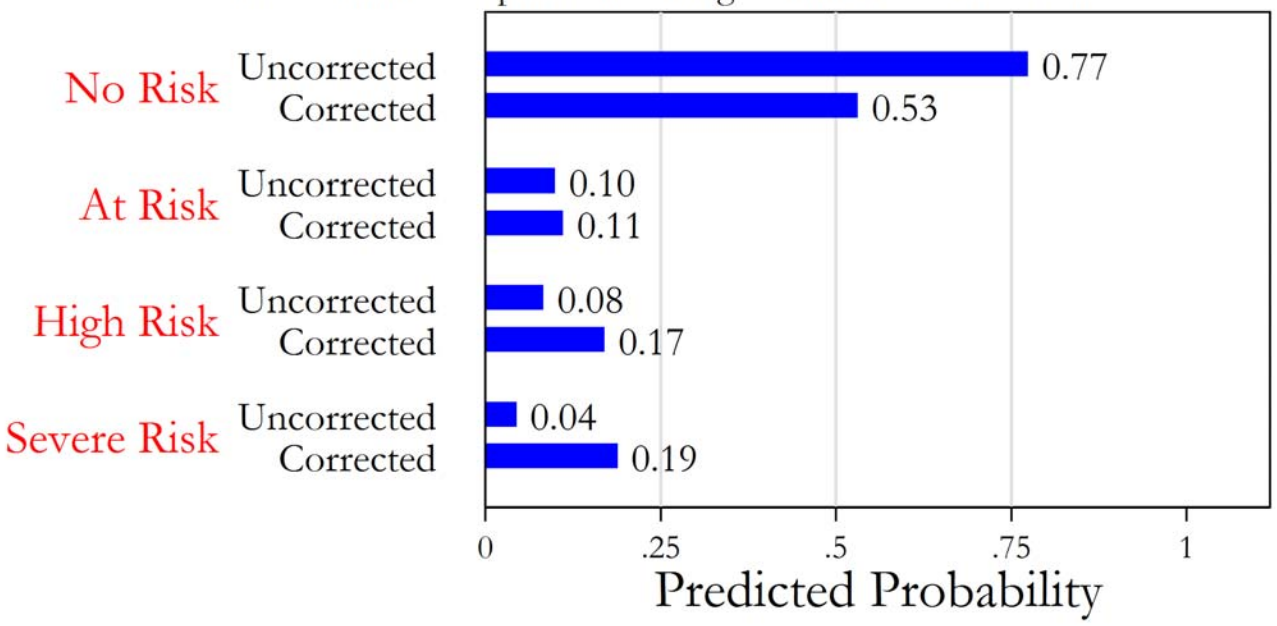


Figure 2: Statistical Significance of Sample Selection Corrections for Nicotine Dependence

100 predicted marginal probabilities from each model, for each individual, reflecting covariance of estimates

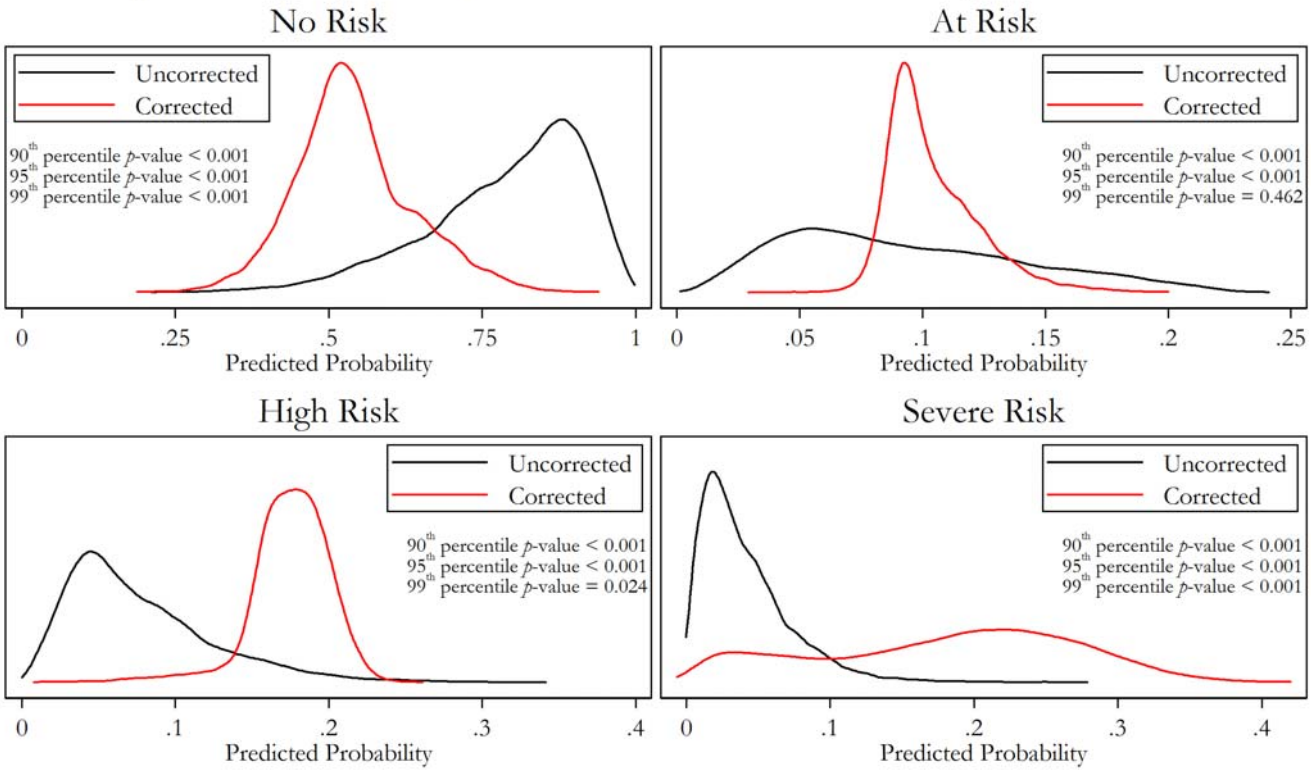


Figure 3: Probability of Current Nicotine Dependence By Years Since Onset of Smoking for Current Daily Smokers

Source: *National Epidemiological Survey on Alcohol & Related Conditions (NESARC)* of 2000/1

Estimated Nicotine Dependence correcting for sample selection

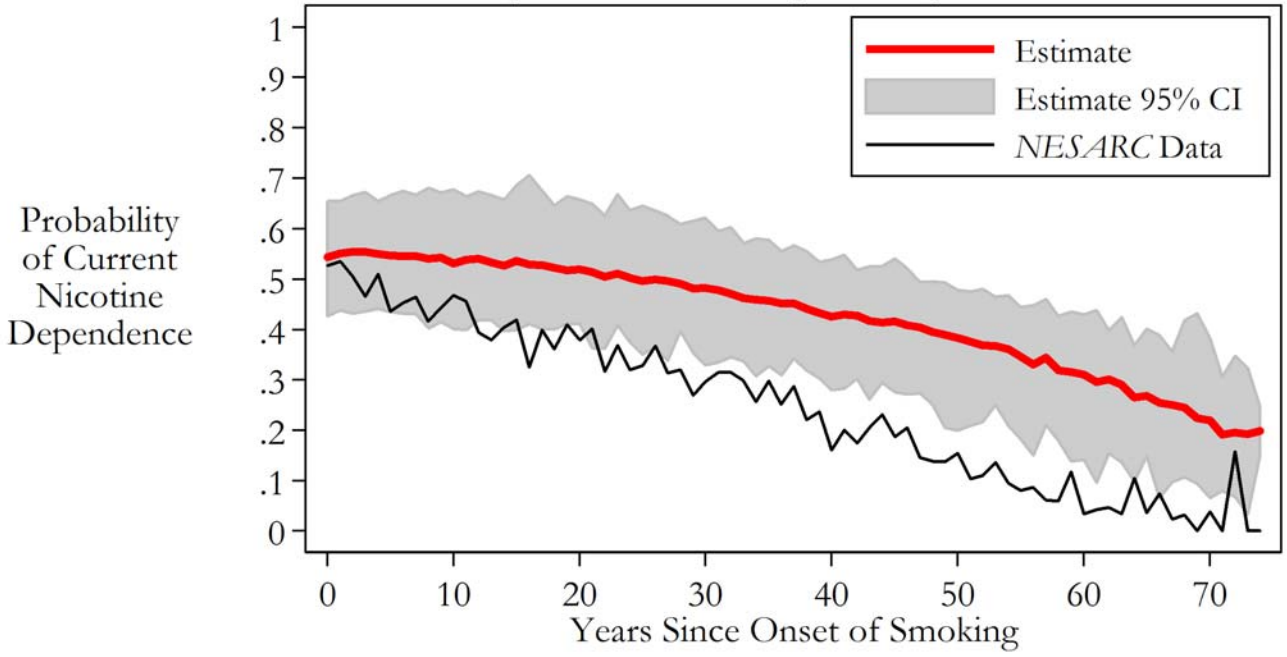
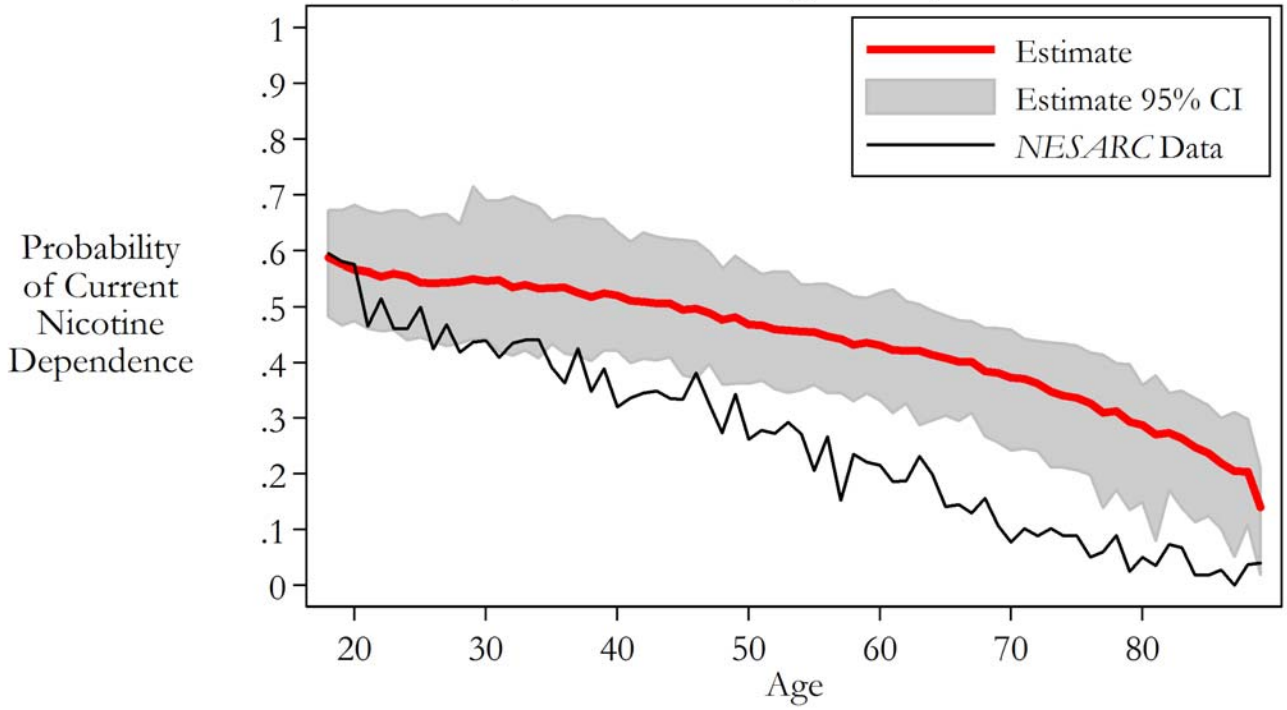


Figure 4: Probability of Current Nicotine Dependence by Age of Current Daily Smoker

Source: *National Epidemiological Survey on Alcohol & Related Conditions (NESARC)* of 2000/1

Estimated Nicotine Dependence correcting for sample selection



References

American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV)* Washington DC: APA Press, 1994.

American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders 5 (DSM 5)* Washington DC: APA Press, 2013.

Bärnighausen, Till; Bor, Jacob; Wandira-Kazibwe, Speciosa, and Canning, David, "Correcting HIV Prevalence Estimates for Survey Nonparticipation Using Heckman-Type Selection Models," *Epidemiology*, 22(1), January 2011a, 27-35.

Bärnighausen, Till; Bor, Jacob; Wandira-Kazibwe, Speciosa, and Canning, David, "Interviewer Identity as Exclusion Restriction in Epidemiology," *Epidemiology*, 22(3), May 2011b, 446.

Cawley, John, and Choi, Anna, "Health Disparities Across Education: the Role of Differential Reporting Error," *IZA Discussion Paper 9141*, June 2015, Institute for the Study of Labor, Bonn.

Chaix, Basile; Bilaudeau, Nathalie; Thomas, Frédérique; Havard, Sabrina; Evans, David; Kestens, Yan, and Bean, Kathy, "Neighborhood Effects on Health: Correcting Bias From Neighborhood Effects on Participation," *Epidemiology*, 22(1), January 2011, 18-26.

Coller, Maribeth; Harrison, Glenn W., and McInnes, Melayne M., "Evaluating the Tobacco Settlement: Are the Damages Awards Too Much or Not Enough?" *American Journal of Public Health*, 92(6), 984-989, 2002.

Derthick, Martha A., *Up In Smoke: From Legislation to Litigation in Tobacco Politics* (Washington, DC: CQ Press, 2011, Third Edition).

De Luca, Giuseppe, "SNP and SML Estimation of Univeriate and Bivariate Binary-Choice Models," *Stata Journal*, 8(2), 2008, 190-220.

De Luca, Giuseppe, and Perotti, Valeria, "Estimation of Ordered Response Models with Sample Selection," *Stata Journal*, 11(2), 2011, 213-239.

Gallant, A. Ronald, and Douglas W. Nychka, "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 55(2), March 1987, 363–390.

Geneletti, Sara; Mason, Alexina, and Best, Nicky, "Commentary: Adjusting for Selection Effects in Epidemiologic Studies: Why Sensitivity Analysis is the Only 'Solution'," *Epidemiology*, 22(1), January 2011, 36-39.

Grant Bridget F.; Dawson, Deborah A.; Stinson, Frederick S.; Chou, S. Patricia; W. Kay, and R. Pickering, "The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): Reliability of Alcohol Consumption, Tobacco Use, Family History of

Depression and Psychiatric Diagnostic Modules in a General Population Sample," *Drug and Alcohol Dependence*, 71, 2003, 7-16.

Grant, Bridget F.; Stinson, Frederick S.; Dawson, Deborah A.; Chou, S. Patricia; Ruan, W. June, and Pickering, Roger P., "Co-occurrence of 12-Month Alcohol and Drug Use Disorders and Personality Disorders in the United States: Results From the National Epidemiologic Survey on Alcohol and Related Conditions," *Archives of General Psychiatry*, 61, April 2004a, 361-368.

Grant, Bridget F.; Hasin, Deborah S.; Chou, S. Patricia; Stinson, Frederick S., and Dawson, Deborah A., "Nicotine Dependence and Psychiatric Disorders in the United States: Results From the National Epidemiologic Survey on Alcohol and Related Conditions," *Archives of General Psychiatry*, 61, November 2004b, 1107-1115.

Gorber, Sarah Connor; Shields, Margot; Tremblay, Mark S, and McDowell, Ian, "Feasibility of establishing correction factors to adjust self-reported estimates of obesity," *Health Reports*, 19(3), September 2008, 71-82, Statistics Canada Catalogue #82-003-XPE.

Gorber, Sarah Conner; Schofield-Hurwitz, Sean; Hardt, Jill; Levasseur, Geneviève, and Tremblay, Mark, "The Accuracy of Self-Reported Smoking: A Systematic Review of the Relationship Between Self-Reported and Cotinine-Assessed Smoking Status," *Nicotine & Tobacco Research*, 11, 2009, 12-24.

Harrison, Glenn W.; Jessen, Lasse J.; Lau, Morten, and Ross, Don, "Disordered Gambling Prevalence: Methodological Innovations in a General Danish Population Survey," *Journal of Gambling Studies*, forthcoming 2017.

Harrison, Glenn W.; Lau, Morten, and Yoo, Hong Il, "Risk Attitudes, Sample Selection and Attrition in a Longitudinal Field Experiment," *CEAR Working Paper 2014-04*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2014.

Harrison, Glenn W.; Lau, Morten I., and Rutström, E. Elisabet, "Risk Attitudes, Randomization to Treatment, and Self-Selection Into Experiments," *Journal of Economic Behavior and Organization*, 70(3), June 2009, 498-507.

Heatherton, Todd F.; Kozlowski, Lyn T.; Frecker, Richard C., and Fagerström, Karl-Olov, "The Fagerström Test for Nicotine Dependence: A Revision of the Fagerström Tolerance Questionnaire," *British Journal of Addiction*, 86, 1991, 1119-1127.

Heckman, James J., "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 1976, 5, 475–492.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), January 1979, 153-162.

Hernán, Miguel A.; Hernández-Diaz, Sonia, and Robins, James M., "A Structural Approach to Selection Bias," *Epidemiology*, 15(5), September 2004, 615-625.

Rothman, Kenneth J.; Greenland, Sander, and Lash, Timothy, L., *Modern Epidemiology* (New York: Lippincott, Williams & Wilkin, 3rd Edition, 2012).

Stewart, Mark B., "Semi-nonparametric Estimation of Extended Ordered Probit Models," *Stata Journal*, 4(1), 2004, 27-39.

Stewart, Mark B., "A Comparison of Semiparametric Estimators for the Ordered Response Model," *Computational Statistics & Data Analysis*, 49, 2005, 555-573.

Appendix (NOT FOR PUBLICATION)

A.1 Reporting Bias and Nicotine

The possibility of sample selectivity because of the use of trigger questions is a particular form of reporting bias. There is a large literature in the area of smoking that addresses reporting bias. All of it focuses on smoking prevalence, which is important. But prevalence is not the same as dependence, which is the focus here, so that evidence is not relevant to the inferential issue considered here. Prevalence includes current and former smokers, and even current smoking is only a necessary condition for someone being nicotine dependent.

The general way in which reporting bias has been examined is by comparing the usual self-reports with physical evidence of exposure to nicotine. One difficulty with these comparisons is that the reference periods may not be the same. The time horizon over which cotinine can be detected in urine or hair samples is a matter of days, whereas many self reports cover much longer periods. Even the expression "current smoker" need not mean that the person is smoking daily, or even in the last few days, in colloquial usage. Another difficulty with these comparisons is that they are unable to be used to address the reliability of self-reports of former or "irregular" smokers. Another difficulty is that cotinine may be present because of nicotine replacement therapy and/or exposure to "secondhand" smoke.

One of the best comparisons comes from the National Health and Nutrition Examination Survey (NHANES), used by Cawley and Choi [2015]. The NHANES conducts surveys and also collects samples from a subset of respondents, and has been doing so every year since 1999.[19] Cawley and Choi [2015] study data between 1999 and 2012 for respondents that have both data, and meet several modest restrictions. One of the NHANES smoking questions asks if the person has smoked (or used any nicotine-related product) in the last 5 days (SMQ.680), which accords relatively closely to the horizon for detection from the serum cotinine detected from a urine sample, which is viewed as between "a few days and up to a week" in clinical settings. The respondent was also asked questions about the intensity of smoking: how many of the last 5 days they smoked (SMQ.710), and the number of cigarettes smoked per day on those days (SMQ.720a). Urinary cotinine is measured in ng/ml by NHANES, and Cawley and Choi [2015] convert this to two binary thresholds of 15ng/ml and 3ng/ml to define an "objective smoker." Clearly there is rich data here to evaluate the relationship between reports and cotonine. Using the self-report of any use in the last 5 days, and the two objective thresholds, Cawley and Choi [2015] show that self-reports become more accurate as someone has greater formal education. The reason for this effect is not so relevant for our purposes, just that there is an effect that varies with a demographic characteristic of the individual. It is exactly this sort of heterogeneity in response bias that is driving selectivity effects. Of course, one issue is that NHANES respondents knew at the outset that there could be some later objective clinical evaluation of their responses, and this could have affected response bias.

Gorber et al. [2009] survey the literature on reporting bias and smoking measured by cotinine. They conclude that there is some evidence of a reporting bias that understates observed measures, roughly between 1 and 4 percentage points, but that bias measures differ significantly in terms of which objective measures are used and which self-reports are used.

[19] Certain aspects of the NHANES were conducted as far back as the 1960s.

A.2 The NESARC Sample Frame and Survey Administration

Grant et al. [2004a; p.362] provide an authoritative description of the sample frame and survey administration for the NESARC data used:

> The Wave 1 NESARC is a nationally representative face-to-face survey of 43093 respondents, aged 18 years and older, conducted by the NIAAA in 2001 through 2002. The target population of the Wave 1 NESARC is the civilian, noninstitutionalized population residing in the United States and the District of Columbia, including Alaska and Hawaii. The housing-unit sampling frame of the NESARC was the US Census Bureau Census 2000 Supplementary Survey, a national survey of more than 78000 households per month conducted in 2000 through 2001. The NESARC also included a group-quarters sampling frame derived from the Census 2000 Group Quarters Inventory. The group-quarters sampling frame captures important subgroups of the population with heavy substance use patterns (eg, college housing) not often included in general population surveys. The sampling frame response rate was 99%, the household response rate was 89%, and the person response rate was 93%, yielding an overall survey response rate of 81%, substantially higher than other surveys of this kind.
>
> Information on race and ethnicity collected in the Census 2000 Supplementary Survey in 2000 through 2001 was used to oversample African American and Hispanic households. The oversampling procedure increased the percentage of non-Hispanic, African American households in the sample from 12.3% to 19.1% (n=8245) and the percentage of Hispanic households from 12.5% to 19.3% (n=8308). One sample person from each household or group-quarters unit was randomly selected for interview, and young adults, aged 18 to 24 years, were oversampled at a rate of 2.25 times that of other members in the household.
>
> The NESARC data were weighted to reflect the probabilities of the selection of primary sampling units (PSUs) within strata and for the selection of housing units within the sample PSUs. The PSUs are mutually exclusive categories of persons or units of interest identified in the first stage of the multistage NESARC sample. The PSUs consisted of geographic units representing the entire United States defined in terms of sociodemographic criteria. The data also were weighted: (1) to account for the selection of 1 sample person from each household; (2) to account for oversampling of young adults; (3) to adjust for nonresponse at the household level and person level; and (4) to reduce the variance arising from selecting 2 PSUs to represent an entire stratum. The weighted data were then adjusted to be representative of the US civilian, noninstitutionalized population for a variety of socioeconomic variables including region, age, sex, race, and ethnicity using the 2000 Decennial Census of Population and Housing and statistics on births, deaths, immigration and emigration, and the size of the Armed Forces.
>
> Approximately 1800 experienced lay interviewers from the US Census Bureau administered the NESARC using laptop computer–assisted software that included built-in skip, logic, and consistency checks. On average, the interviewers had 5 years' experience working on census and other health-related national surveys. All NESARC

interviewers completed a 5-day selfstudy at home and participated in a standardized 5-day inclass training session at 1 of the bureau's 12 regional offices. The NESARC training supervisors from each regional office also were required to complete the home study and to attend a centralized training session prior to fielding of the survey, where they completed the in-class training under the direction of NIAAA sponsors and Census Field and Demographics Survey Division headquarters staff.

Regional supervisors recontacted a random 10% of all respondents for quality-control purposes. In these quality control interviews, a series of questions were reasked to verify that respondents had received the entire interview and that the questionnaire had been administered properly. There was no case in which it was determined that the interview had been conducted in any manner that was inconsistent with the interviewer's extensive training. In addition, 2657 respondents were randomly selected to participate in a reinterview study after completion of their NESARC interview. Each respondent was readministered 1 to 3 sections of the survey assessment instrument. These interviews not only served as an additional check on survey data quality but formed the basis of a test-retest reliability study of new modules of the survey instrument.

A.3 Survey Questions on Nicotine Dependence

Current ND in the past 12 months is defined in the DSM-IV by the presence of 3 or more of 7 criteria within the past 12 months, for *someone that is currently a daily smoker*. This last qualification is important, and natural: someone cannot begin to be classified as currently nicotine dependent unless they are actually smoking daily. The NESARC contained a number of questions that mapped into these criteria. Each question is listed below, in slightly paraphrased form. If any question within the group of questions for a specific criterion was answered affirmatively, the criterion in bold was considered to be satisfied:

1. **Tolerance**
    a. Find that you had to use much more tobacco that you once did to get the effect you wanted?
    b. Increase your use of tobacco by at least 50 percent?
2. **Withdrawal**
    a. Wake up in the middle of the night to use tobacco?
    b. Often use tobacco just after getting up or shortly after getting up in the morning?
    c. Find yourself using tobacco just after being in a situation where tobacco use was not permitted – like after being on a plane, at a meeting, or shopping at the mall?
    d. Four or more of the following symptoms:
        i. Depressed
        ii. Difficulty falling asleep or staying asleep
        iii. Difficulty concentrating
        iv. Eat more than usual or gain weight
        v. Become easily irritated, angry, or frustrated
        vi. Feel anxious or nervous

vii. Feel your heart beating more slowly than usual
viii. Feel more restless than usual

e. Did you use tobacco to keep from having any of these experiences? (referring to part d of criterion #2, immediately above)

3. **Larger amount/longer period**
   a. Have a period when you often used tobacco more than you intended to?
4. **Difficulty cutting down**
   a. More than once try to stop or cut down on your tobacco use but found you couldn't do it?
   b. More than once want to stop or cut down on your tobacco use?
5. **Activities to obtain, use, recover**
   a. Find yourself (chain smoking/using one pinch or plug of snuff or chewing tobacco right after another)?
6. **Reduce other activities**
   a. Give up or cut down on activities that you were interested in or that gave you pleasure because tobacco use was not permitted at the activity?
   b. Give up or cut down on activities that were important to you – like associating with friends or relatives or attending social activities because tobacco use was not permitted at the activity?
7. **Use despite problems**
   a. Continue to use tobacco even though you knew it was causing you a health problem or making a health problem worse?
   b. Continue to use tobacco even though it made you nervous, jittery, anxious or depressed.

The formal classification of ND in *DSM-IV* is usually binary: if someone meets the threshold of being a current daily smoker and "ticks off" 3 or more of these 7 criteria, then they are classified as being ND. A similar set of criteria are used to define Tobacco Dependence in the 10th Revision of the International Classification of Diseases (ICD-10).

If someone meets the criteria for current ND, *lifetime* ND can also be defined by asking the same questions about the period before the last 12 months, and in addition confirming that one or three criteria are satisfied for the month [X] one year prior to interview:

1. Before last [X] was there ever a period when some of these experiences were happening around the same time most days for at least a month?
2. Before last [X] was there ever a period when some of these experiences were happening around the same time on and off for a few months or longer?
3. Before last [X] was there ever a period when some of these experiences happened within the same 1-year period?

A.4 Background Questions

The NESARC questionnaire contains a number of background questions asked of every respondent prior to any of the trigger questions. The following battery of questions are used to generate

21 binary variables:

```
Variable Description
-------------------------------------------------------------------------------
ss_1     Sample Selection variable 1 from Screening Question #18 Part 1
ss_2     Sample Selection variable 2 from Screening Question #18 Part 2
ss_3     Sample Selection variable 3 from Screening Question #19 Part 1
ss_4     Sample Selection variable 4 from Screening Question #19 Part 2
ss_5     Sample Selection variable 5 from Screening Question #20
ss_6     Sample Selection variable 6 from Screening Question #21 Part 1
ss_7     Sample Selection variable 7 from Screening Question #21 Part 2
ss_8     Sample Selection variable 8 from Screening Question #21 Part 3
ss_9     Sample Selection variable 9 from Screening Question #22
ss_10    Sample Selection variable 10 from Screening Question #23 Part 1
ss_11    Sample Selection variable 11 from Screening Question #23 Part 2
ss_12    Sample Selection variable 12 from Screening Question #23 Part 3
ss_13    Sample Selection variable 13 from Screening Question #23 Part 4
ss_14    Sample Selection variable 14 from Screening Question #23 Part 5
ss_15    Sample Selection variable 15 from Screening Question #23 Part 6
ss_16    Sample Selection variable 16 from Screening Question #23 Part 7
ss_17    Sample Selection variable 17 from Screening Question #23 Part 8
ss_18    Sample Selection variable 18 from Screening Question #23 Part 9
ss_19    Sample Selection variable 19 from Screening Question #23 Part 10
ss_20    Sample Selection variable 20 from Screening Question #23 Part 11
ss_21    Sample Selection variable 21 from Screening Question #23 Part 12
```

The original NESARC questions referenced here are as follows:

Section 1 - BACKGROUND INFORMATION (Continued)

(SHOW FLASHCARD 11C)

18. **During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as the result of your physical health?**

(1) **Accomplished less than you would like.**

1 ☐ All of the time
2 ☐ Most of the time
3 ☐ Some of the time
4 ☐ A little of the time
5 ☐ None of the time

(2) **Were limited in the kind of work or other activities.**

1 ☐ All of the time
2 ☐ Most of the time
3 ☐ Some of the time
4 ☐ A little of the time
5 ☐ None of the time

(SHOW FLASHCARD 11C)

19. **During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as the result of any emotional problems such as feeling depressed or anxious?**

(1) **Accomplished less than you would like.**

1 ☐ All of the time
2 ☐ Most of the time
3 ☐ Some of the time
4 ☐ A little of the time
5 ☐ None of the time

(2) **Didn't do work or other activities as carefully as usual.**

1 ☐ All of the time
2 ☐ Most of the time
3 ☐ Some of the time
4 ☐ A little of the time
5 ☐ None of the time

(SHOW FLASHCARD 11D) 20. **During the past 4 weeks, how much did pain interfere with your normal work including both work outside the home and housework?**	1 ☐ Not at all 2 ☐ A little bit 3 ☐ Moderately 4 ☐ Quite a bit 5 ☐ Extremely
(SHOW FLASHCARD 11C) 21. **The next few questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks...**	
(1) **Have you felt calm and peaceful?**	1 ☐ All of the time 2 ☐ Most of the time 3 ☐ Some of the time 4 ☐ A little of the time 5 ☐ None of the time
(2) **Did you have a lot of energy?**	1 ☐ All of the time 2 ☐ Most of the time 3 ☐ Some of the time 4 ☐ A little of the time 5 ☐ None of the time
(3) **Have you felt downhearted and depressed?**	1 ☐ All of the time 2 ☐ Most of the time 3 ☐ Some of the time 4 ☐ A little of the time 5 ☐ None of the time
(SHOW FLASHCARD 11C) 22. **During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities like visiting with friends, relatives, and so forth?**	1 ☐ All of the time 2 ☐ Most of the time 3 ☐ Some of the time 4 ☐ A little of the time 5 ☐ None of the time

23. **Can you please tell me if you have had any of the following experiences in the last 12 months?** **In the last 12 months. . .** *(Repeat phraze frequently)*	
(1) **Did any of your family members or close friends die?**	1 ☐ Yes 2 ☐ No
(2) **Did any of your family members or close friends have a serious illness or injury?**	1 ☐ Yes 2 ☐ No
(3) **Did you move or have anyone new come to live with you?**	1 ☐ Yes 2 ☐ No
(4) **Were you fired or laid off from a job?**	1 ☐ Yes 2 ☐ No
(5) **Were you unemployed and looking for a job for more than a month?**	1 ☐ Yes 2 ☐ No
(6) **Have you had trouble with your boss or a coworker?**	1 ☐ Yes 2 ☐ No
(7) **Did you change jobs, job responsibilities or work hours?**	1 ☐ Yes 2 ☐ No
(8) **Did you get separated or divorced or break off a steady relationship?**	1 ☐ Yes 2 ☐ No
(9) **Have you had serious problems with a neighbor, friend or relative?**	1 ☐ Yes 2 ☐ No
(10) **Have you experienced a major financial crisis, declared bankruptcy or more than once been unable to pay your bills on time?**	1 ☐ Yes 2 ☐ No
(11) **Did you or a family member have trouble with the police, get arrested or get sent to jail?**	1 ☐ Yes 2 ☐ No
(12) **Were you or a family member the victim of any type of crime?**	1 ☐ Yes 2 ☐ No

Binary variables numbered 1, 2, 3, 4, 8 and 9 are defined as 1 if the NESARC question response is not #5, variables numbered 5, 6 and 7 are defined as 1 if the NESARC question response is not #1, and the remaining variables are defined as 1 if the NESARC question 23 response is #1.

A.5 Formal Statistical Specification

The statistical model is an extension of the standard Heckman model in three ways:

- The main equation is an ordered probit instead of OLS specification
- The bivariate distribution between the selection equation and the main equation is specified in a semi-nonparament (SNP) manner.

The formal specification is presented in equations (1), (2) and (3) of the main text. These are repeated here for convenience:

$$Y_j^* = \beta_j^\top X_j + U_j \qquad j = 1, 2 \qquad (1)$$

$$Y_1 = \mathbf{I}(Y_1^* \geq 0) \qquad (2)$$

$$Y_2 = \Sigma_{h=0...H}\, h\, \mathbf{I}(\alpha_h < Y_2^* \leq \alpha_{h+1}) \qquad \text{if } Y_1 = 1 \qquad (3)$$

Notation is defined in the main text.

There are four identifiability restrictions on this model. The first is familiar from the parametric ordered probit and ordered logit models, that the intercept in β_2 be set to zero. The second is that the exogenous variables X_1 contain at least one variable not contained in X_2. Although this "exclusion restriction" is not formally needed in the parametric model, identification is then only likely to be weak (Meng and Schmidt [1985] and Keane [1992]). However, this restriction is formally needed in the semi-parametric case (Lee [1995]). In our case we have a long list of variables that meet this exclusion restriction, shown later in Tables A1 and A2. Third, X_1 and X_2 must each contain one continuous variable, or have sufficient coverage from a rich list of discrete variables (Manski [1988]). In our case we have variables denoting age (and the square of age) and the weight of the individual, as well as rich array of discrete variables. Fourth, since the means of U_1 and U_2 are not constrained to be zero with the SNP estimator, the intercept of β_1 and the first threshold α_1 are set equal to the parametric estimates.

The estimation of this SNP ordered probit model is based on software developed by De Luca and Perotti [2011], which in turn is based on components developed by Stewart [2005] and De Luca [2008]. All computer programs to process the raw NESARC data, estimate the model, generate marginal effects and Figures 1-4 will be available at http://cear.gsu.edu/gwh/. The software runs on *Stata* version 15, although all estimations run on *Stata* version 14.

A.6 Estimates

Tables A1 through A12 document the definition of the variables used in the empirical model (Table A1), descriptive statistics of the sample (Table A2), the estimates of the SNP Ordered Probit model (Table A3), marginal effects for each covariate with respect to the selection probability (Table A4), marginal effects for each covariate with respect to marginal (unconditional) probabilities of each risk level of nicotine dependence (Tables A5 through A8), and marginal effects for each covariate with

respect to conditional probabilities of each risk level of nicotine dependence given selection (Tables A9 through A12). All marginal effects are *average* marginal effects, calculated at the actual values of covariates for each respondent and then averaged.[20]

Table A1 is a glossary of all covariates. Most are self-explanatory, and the ones that reference specific screening questions from the NESARC are documented in §A.4. We see that there are many variables in the sample selection equation that are not in the main outcome equation. Table A2 provides descriptive statistics, and we see that there are several non-binary, "roughly continuous" variables in both equations.

Table A3 lists estimates of the SNP Ordered Probit model. In the body of the table are the estimates for the covariates of the main equation for nicotine disorder risk level, the selection equation, and then the two estimated thresholds α_2 and α_3, referred to as Threshold #2 and Threshold #3, respectively. The estimate for α_1 is fixed at the estimate from the parametric Ordered Probit model, and listed as -1.708 in the first subtitle text. There are also R×R polynomial coefficients to be estimated, but these are by themselves of no interest: what is of interest is the implied joint distribution of the errors U_1 and U_2. These are summarized in terms of their mean, standard deviation, skewness and kurtosis in the second block of subtitle text, as well as their correlation $\varrho = -0.793$. The correct parametric alternative would have a mean of 0, skewness of 0 and kurtosis of 3 for each error.[21] Evaluating the log-likelihood of the SNP Ordered Probit model at those parametric values produces a value of only -120149, and the unrestricted SNP Ordered Probit model has a log-likelihood of -73027. A χ^2 test of these log-likelihoods, with R×R-2 degrees of freedom, has a *p*-value less than 0.001, leading us to reject the null hypothesis that the SNP Ordered Probit provides the same estimates as the parametric Ordered Probit.

The individual coefficient estimates of the covariates in Table A3 are of less direct interest than the implied marginal effects, evaluated in Tables A4 through A12. However, it is apparent that many are statistically significant, in both equations.

Table A4 shows the calculated marginal effects of each covariate on the probability of being selected to be asked the nicotine dependence DSM questions. We see, for instance, that women have a 0.032 lower probability of being selected into the DSM questions than men, and this effect is statistically significant at less than a 0.001 level. Similarly, those with a college or graduate education are 0.03 and 0.06 *less* likely to be selected, respectively, whereas those with incomes below $20,000 and between $20,000 and $35,000 are 0.06 and 0.07 *more* likely to be selected, respectively. If the correlation between the errors of the selection equation and nicotine dependence equation had been zero, these patterns of heterogeneity in the selection process would have had no effect on inferences about nicotine dependence.

Table A5 shows the marginal effects of each covariate on the unconditional marginal probability of being classified as having No Risk for nicotine dependence. This unconditional probability takes into account the marginal effect of being selected, and then the marginal effect of being classified as having

---

[20] An alternative is often to calculate the marginal effects at the sample mean values of all covariates.

[21] The word "correct" is added, because the identifying restrictions for the parametric Ordered Probit model differ from the identifying restrictions for the SNP counterpart. The former include constraining the variance of U_1 and U_2 each to 1. The correct parametric counterpart for the purpose of the likelihood ratio test is the SNP Ordered Probit model evaluated at the starting values developed for the model.

No Risk conditional on being selected, which is shown in Table A9. Table A5 shows that women are 0.025 more likely to be classified as having No Risk, and Table A9 shows that women are 0.057 more likely to be classified as having No Risk if selected, and both marginal effects are statistically significant. Coupled with the marginal effect for women in the selection process, this is an instance in which all data-generating processes modeled here are pointing in the same qualitative direction: not only are women less likely to be selected to be asked the DSM questions, even if they are selected they are more likely to be classified as having No Risk.

As a general matter the *qualitative* pattern illustrated for women applies to all covariates, whether or not it is in one direction or the other. What matters for sample selection is the *strength* of the effects across covariates, given that there is a non-zero correlation between the errors. If the prevalence of DSM-IV Nicotine Dependence increases from 12.7% to 35.8%, as shown in Figure 1, then the process driving that result comes from the marginal effects of those *more* likely to be selected and *more* likely to be classified as at risk being greater, when weighted by sample size for that covariate, than the marginal effects of those *less* likely to be selected and *less* likely to be classified as at risk.

Additional References

Keane, Michael P., "A note on identification in the multinomial probit model," *Journal of Business and Economic Statistics*, 10, 1992, 193-200.

Lee, Lung-Fei, "Semiparametric maximum likelihood estimation of polychotomous and sequential choice models," *Journal of Econometrics*, 65, 1995, 381-428.

Manski, Charles F., "Identification of binary response models," *Journal of the American Statistical Association*, 83, 1988, 729-738.

Meng, Chun-Lo, and Schmidt, Peter, "On the cost of partial observability in the bivariate probit model," *International Economic Review*, 26, 1985, 71-85.

Wong, Suzy L.; Shields, Margot; Leatherdale, Scott; Malaison, Eric, and Hammon, David, "Assessment of validity of self-reported smoking status," *Health Reports*, 23(1), March 2012, 1-7 (Statistics Canada, Catalogue no. 82-003-XPE).

Table A1: Glossary of Covariates

Variable	Definition
	Main Disorder Equation
female	Female
black	Black or Afro-American
hispanic	Hispanic
age	Age in years
age2	Age in years (squared)
separated	Separated by being widowed, divorced or separated
midwest	Midwest region
south	South region
west	West region
high_school	Completed High School or GED
some_college	Completed some college
college	Completed a college degree
graduate	Completed a graduate degree
income1	Personal income below $20,000
income2	Personal income between $20,000 and $35,000
income4	Personal income of $70,000 or more
welfare	Receives some form of welfare
	Additional Variables for Selection Equation
height	Height in feet
weigh	Weight in stones
friday	Interview conducted on a Friday
saturday	Interview conducted on a Saturday
sunday	Interview conducted on a Sunday
personality_battery1	Responses to questions on usual feelings and actions
personality_battery2	Responses to questions on effects of feelings and actions
ss_1	Sample Selection variable 1 from Screening Question #18 Part 1
ss_2	Sample Selection variable 2 from Screening Question #18 Part 2
ss_3	Sample Selection variable 3 from Screening Question #19 Part 1
ss_4	Sample Selection variable 4 from Screening Question #19 Part 2
ss_5	Sample Selection variable 5 from Screening Question #20
ss_6	Sample Selection variable 6 from Screening Question #21 Part 1
ss_7	Sample Selection variable 7 from Screening Question #21 Part 2
ss_8	Sample Selection variable 8 from Screening Question #21 Part 3
ss_9	Sample Selection variable 9 from Screening Question #22
ss_10	Sample Selection variable 10 from Screening Question #23 Part 1
ss_11	Sample Selection variable 11 from Screening Question #23 Part 2
ss_12	Sample Selection variable 12 from Screening Question #23 Part 3
ss_13	Sample Selection variable 13 from Screening Question #23 Part 4
ss_14	Sample Selection variable 14 from Screening Question #23 Part 5
ss_15	Sample Selection variable 15 from Screening Question #23 Part 6
ss_16	Sample Selection variable 16 from Screening Question #23 Part 7
ss_17	Sample Selection variable 17 from Screening Question #23 Part 8
ss_18	Sample Selection variable 18 from Screening Question #23 Part 9
ss_19	Sample Selection variable 19 from Screening Question #23 Part 10
ss_20	Sample Selection variable 20 from Screening Question #23 Part 11
ss_21	Sample Selection variable 21 from Screening Question #23 Part 12

Table A2: Descriptive Statistics for Covariates

Variable	N	Average	Standard Deviation	Minimum	Maximum
Main Disorder Equation					
female	43093	.5702782	.4950421	0	1
black	43093	.1995684	.3996806	0	1
hispanic	43093	.1927923	.394496	0	1
age	43093	46.40081	18.17861	18	98
age2	43093	2483.489	1865.155	324	9604
separated	43093	.2579769	.4375263	0	1
midwest	43093	.2086418	.4063425	0	1
south	43093	.3749101	.4841053	0	1
west	43093	.2259532	.4182133	0	1
high_school	43093	.291161	.4543028	0	1
some_college	43093	.2063212	.4046685	0	1
college	43093	.2093844	.4068739	0	1
graduate	43093	.1109925	.3141265	0	1
income1	43093	.4896619	.4998989	0	1
income2	43093	.2314761	.4217808	0	1
income4	43093	.2906969	.4540892	0	1
welfare	43093	.2961502	.4565634	0	1
Additional Variables for Selection Equation					
height	43093	5.549293	.3356052	4	7
weigh	43093	12.17406	2.914927	4.428571	35.71429
friday	43093	.1413919	.3484294	0	1
saturday	43093	.1438981	.3509904	0	1
sunday	43093	.0768106	.2662938	0	1
personality_battery1	43093	3.466642	4.772873	0	55
personality_battery2	43093	3.443552	4.769596	0	55
ss_1	43093	.348154	.4763907	0	1
ss_2	43093	.3008841	.4586478	0	1
ss_3	43093	.2764718	.4472581	0	1
ss_4	43093	.2422435	.4284459	0	1
ss_5	43093	.3640963	.4811814	0	1
ss_6	43093	.7625601	.425519	0	1
ss_7	43093	.7989697	.4007753	0	1
ss_8	43093	.4865291	.4998243	0	1
ss_9	43093	.2665166	.4421425	0	1
ss_10	43093	.3197503	.4663851	0	1
ss_11	43093	.3476435	.4762276	0	1
ss_12	43093	.1458938	.3530038	0	1
ss_13	43093	.0624463	.2419672	0	1
ss_14	43093	.0896897	.2857401	0	1
ss_15	43093	.079224	.2700912	0	1
ss_16	43093	.2083633	.4061427	0	1
ss_17	43093	.0647437	.2460759	0	1
ss_18	43093	.0561576	.2302285	0	1
ss_19	43093	.1168867	.321289	0	1
ss_20	43093	.0552526	.2284753	0	1
ss_21	43093	.0641636	.2450469	0	1

Table A3: Semi-Nonparametric Ordered Logit Estimates for Nicotine Dependence

Threshold #1 fixed at Ordered Probit estimate: -1.708

Estimated moments of error distributions

Disorder equation	Selection equation
Mean: -0.021	Mean: -0.066
Standard deviation: 1.726	Standard deviation: 1.846
Skewness: -0.159	Skewness: 0.142
Kurtosis: 2.011	Kurtosis: 1.956

Estimated correlation coefficient: -0.793

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Disorder		
female	-0.175	<0.001	-0.244	-0.106
black	-0.225	<0.001	-0.301	-0.148
hispanic	-0.563	<0.001	-0.655	-0.472
age	0.022	<0.001	0.013	0.032
age2	-0.001	<0.001	-0.001	-0.001
separated	0.422	<0.001	0.342	0.503
midwest	0.094	0.043	0.003	0.184
south	-0.009	0.828	-0.092	0.073
west	-0.121	0.010	-0.212	-0.029
high_school	-0.214	<0.001	-0.301	-0.126
some_college	-0.359	<0.001	-0.456	-0.263
college	-0.540	<0.001	-0.649	-0.431
graduate	-0.847	<0.001	-0.993	-0.701
income1	-0.783	<0.001	-0.966	-0.600
income2	-0.769	<0.001	-0.959	-0.578
income4	-0.902	<0.001	-1.098	-0.706
welfare	0.111	0.011	0.025	0.197
		Selection		
female	-0.213	<0.001	-0.270	-0.155
black	-0.212	<0.001	-0.270	-0.154
hispanic	-0.237	<0.001	-0.306	-0.169
age	0.055	<0.001	0.048	0.062
age2	-0.000	<0.001	-0.001	-0.000
separated	0.040	0.150	-0.014	0.094
midwest	-0.011	0.731	-0.076	0.053
south	-0.049	0.103	-0.107	0.010
west	-0.107	0.001	-0.172	-0.041
high_school	0.004	0.900	-0.061	0.069
some_college	0.013	0.727	-0.059	0.084
college	-0.185	<0.001	-0.265	-0.106
graduate	-0.381	<0.001	-0.484	-0.278
income1	0.371	<0.001	0.223	0.518
income2	0.403	<0.001	0.261	0.545
income4	0.333	<0.001	0.188	0.478
welfare	0.151	<0.001	0.088	0.214
height	0.422	<0.001	0.386	0.458
weigh	-0.038	<0.001	-0.045	-0.031
friday	-0.019	0.412	-0.065	0.027

saturday	0.035	0.128	-0.010	0.081
sunday	-0.010	0.746	-0.070	0.050
personality_battery1	0.141	<0.001	0.083	0.199
personality_battery2	-0.120	<0.001	-0.178	-0.061
ss_1	0.055	0.050	-0.000	0.110
ss_2	0.045	0.144	-0.015	0.105
ss_3	0.053	0.051	-0.000	0.106
ss_4	-0.062	0.026	-0.117	-0.008
ss_5	0.054	0.010	0.013	0.095
ss_6	0.041	0.072	-0.004	0.086
ss_7	0.124	<0.001	0.076	0.172
ss_8	0.091	<0.001	0.052	0.130
ss_9	-0.000	0.994	-0.045	0.044
ss_10	0.070	<0.001	0.033	0.106
ss_11	0.070	<0.001	0.034	0.107
ss_12	0.184	<0.001	0.137	0.230
ss_13	0.045	0.206	-0.025	0.116
ss_14	0.122	<0.001	0.059	0.184
ss_15	0.114	<0.001	0.056	0.173
ss_16	0.173	<0.001	0.129	0.218
ss_17	0.179	<0.001	0.115	0.244
ss_18	0.130	<0.001	0.063	0.196
ss_19	0.294	<0.001	0.240	0.347
ss_20	0.265	<0.001	0.197	0.332
ss_21	0.112	<0.001	0.049	0.175
Thresholds for Ordered Categories				
Threshold #2	-0.913	<0.001	-0.967	-0.859
Threshold #3	0.218	<0.001	0.101	0.335

Table A4: Marginal Effects for Selection Probability

Marginal effects are *average* marginal effects, evaluated at observed covariates for each respondent

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
female	-0.032	<0.001	-0.040	-0.024
black	-0.034	<0.001	-0.042	-0.026
hispanic	-0.038	<0.001	-0.048	-0.028
age	0.008	<0.001	0.007	0.010
age2	-0.000	<0.001	-0.000	-0.000
separated	0.009	0.046	0.000	0.017
midwest	-0.000	0.934	-0.010	0.009
south	-0.007	0.128	-0.016	0.002
west	-0.017	<0.001	-0.027	-0.007
high_school	0.001	0.902	-0.009	0.010
some_college	0.002	0.759	-0.009	0.013
college	-0.030	<0.001	-0.042	-0.018
graduate	-0.062	<0.001	-0.077	-0.047
income1	0.064	<0.001	0.041	0.087
income2	0.068	<0.001	0.047	0.090
income4	0.057	<0.001	0.035	0.079
welfare	0.024	<0.001	0.015	0.034
height	0.067	<0.001	0.058	0.076
weigh	-0.006	<0.001	-0.007	-0.005
friday	-0.003	0.443	-0.010	0.004
saturday	0.005	0.131	-0.002	0.012
sunday	-0.002	0.733	-0.011	0.008
personality_battery1	0.021	<0.001	0.013	0.030
personality_battery2	-0.018	<0.001	-0.027	-0.009
ss_1	0.008	0.052	-0.000	0.017
ss_2	0.007	0.126	-0.002	0.016
ss_3	0.008	0.054	-0.000	0.016
ss_4	-0.010	0.023	-0.018	-0.001
ss_5	0.008	0.009	0.002	0.015
ss_6	0.006	0.069	-0.001	0.013
ss_7	0.019	<0.001	0.012	0.026
ss_8	0.014	<0.001	0.008	0.020
ss_9	-0.000	0.977	-0.007	0.007
ss_10	0.011	<0.001	0.005	0.017
ss_11	0.011	<0.001	0.005	0.016
ss_12	0.028	<0.001	0.021	0.035
ss_13	0.007	0.193	-0.004	0.018
ss_14	0.019	<0.001	0.009	0.028
ss_15	0.017	<0.001	0.009	0.026
ss_16	0.026	<0.001	0.020	0.033
ss_17	0.027	<0.001	0.017	0.037
ss_18	0.019	<0.001	0.009	0.030
ss_19	0.045	<0.001	0.037	0.053
ss_20	0.041	<0.001	0.031	0.051
ss_21	0.017	<0.001	0.007	0.027

Table A5: Marginal Effects for No Risk Marginal Probability

Unconditional (marginal) probability of No Risk of Nicotine Dependence

Marginal effects are *average* marginal effects, evaluated at observed covariates for each respondent

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
female	0.0255	<0.001	0.0158	0.0353
black	0.0311	<0.001	0.0206	0.0416
hispanic	0.0793	<0.001	0.0658	0.0928
age	-0.0034	<0.001	-0.0048	-0.0019
age2	0.0001	<0.001	0.0001	0.0001
separated	-0.0569	<0.001	-0.0681	-0.0457
midwest	-0.0112	0.0731	-0.0234	0.0010
south	0.0026	0.6490	-0.0087	0.0139
west	0.0168	0.0101	0.0040	0.0296
high_school	0.0314	<0.001	0.0194	0.0434
some_college	0.0527	<0.001	0.0394	0.0659
college	0.0773	<0.001	0.0615	0.0931
graduate	0.1189	<0.001	0.0960	0.1419
income1	0.1326	<0.001	0.1011	0.1640
income2	0.1292	<0.001	0.0980	0.1604
income4	0.1483	<0.001	0.1161	0.1804
welfare	-0.0137	0.0215	-0.0254	-0.0020

Table A6: Marginal Effects for Mild Risk Marginal Probability

Unconditional (marginal) probability of Mild Risk of Nicotine Dependence

Marginal effects are *average* marginal effects, evaluated at observed covariates for each respondent

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
female	-0.00057	0.71236	-0.00358	0.00245
black	-0.00069	0.71132	-0.00435	0.00297
hispanic	-0.00176	0.71091	-0.01108	0.00755
age	0.00007	0.70953	-0.00032	0.00047
age2	-0.00000	0.70949	-0.00001	0.00001
separated	0.00126	0.71189	-0.00545	0.00797
midwest	0.00025	0.72061	-0.00111	0.00161
south	-0.00006	0.76145	-0.00044	0.00032
west	-0.00037	0.70997	-0.00234	0.00159
high_school	-0.00070	0.71148	-0.00439	0.00300
some_college	-0.00117	0.71162	-0.00737	0.00503
college	-0.00172	0.71068	-0.01079	0.00736
graduate	-0.00264	0.71012	-0.01657	0.01129
income1	-0.00294	0.70828	-0.01837	0.01248
income2	-0.00287	0.70895	-0.01794	0.01220
income4	-0.00329	0.70926	-0.02061	0.01402
welfare	0.00030	0.71289	-0.00132	0.00193

Table A7: Marginal Effects for Moderate Risk Marginal Probability

Unconditional (marginal) probability of Moderate Risk of Nicotine Dependence

Marginal effects are *average* marginal effects, evaluated at observed covariates for each respondent

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
female	0.00147	0.22135	-0.00088	0.00382
black	0.00179	0.21756	-0.00106	0.00463
hispanic	0.00456	0.21223	-0.00260	0.01172
age	-0.00019	0.21893	-0.00050	0.00012
age2	0.00001	0.21007	-0.00000	0.00001
separated	-0.00327	0.21952	-0.00850	0.00195
midwest	-0.00064	0.32353	-0.00192	0.00063
south	0.00015	0.65890	-0.00052	0.00082
west	0.00097	0.23873	-0.00064	0.00257
high_school	0.00180	0.22334	-0.00110	0.00471
some_college	0.00303	0.22126	-0.00182	0.00788
college	0.00445	0.21279	-0.00255	0.01144
graduate	0.00684	0.20662	-0.00377	0.01744
income1	0.00762	0.20190	-0.00408	0.01932
income2	0.00743	0.20496	-0.00406	0.01891
income4	0.00852	0.20622	-0.00469	0.02174
welfare	-0.00079	0.26431	-0.00217	0.00060

Table A8: Marginal Effects for Severe Risk Marginal Probability

Unconditional (marginal) probability of Severe Risk of Nicotine Dependence

Marginal effects are *average* marginal effects, evaluated at observed covariates for each respondent

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
female	-0.02642	<0.001	-0.03634	-0.01651
black	-0.03223	<0.001	-0.04305	-0.02141
hispanic	-0.08211	<0.001	-0.09610	-0.06813
age	0.00349	<0.001	0.00195	0.00503
age2	-0.00009	<0.001	-0.00011	-0.00007
separated	0.05893	<0.001	0.04731	0.07054
midwest	0.01156	0.07289	-0.00107	0.02419
south	-0.00272	0.64930	-0.01445	0.00901
west	-0.01739	0.01021	-0.03065	-0.00412
high_school	-0.03250	<0.001	-0.04497	-0.02003
some_college	-0.05454	<0.001	-0.06842	-0.04065
college	-0.08006	<0.001	-0.09663	-0.06350
graduate	-0.12310	<0.001	-0.14682	-0.09939
income1	-0.13723	<0.001	-0.17098	-0.10349
income2	-0.13377	<0.001	-0.16696	-0.10058
income4	-0.15352	<0.001	-0.18782	-0.11922
welfare	0.01420	0.02166	0.00208	0.02631

Table A9: Marginal Effects for No Risk Conditional Probability

Conditional probability of No Risk of Nicotine Dependence given selection

Marginal effects are *average* marginal effects, evaluated at observed covariates for each respondent

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
female	0.0574	<0.001	0.0491	0.0658
black	0.0652	<0.001	0.0561	0.0743
hispanic	0.1219	<0.001	0.1112	0.1326
age	-0.0114	<0.001	-0.0127	-0.0101
age2	0.0002	<0.001	0.0002	0.0002
separated	-0.0706	<0.001	-0.0799	-0.0612
midwest	-0.0119	0.0289	-0.0226	-0.0012
south	0.0091	0.0689	-0.0007	0.0188
west	0.0339	<0.001	0.0228	0.0449
high_school	0.0340	<0.001	0.0237	0.0444
some_college	0.0565	<0.001	0.0450	0.0681
college	0.1121	<0.001	0.0991	0.1252
graduate	0.1871	<0.001	0.1689	0.2053
income1	0.0881	<0.001	0.0580	0.1183
income2	0.0806	<0.001	0.0504	0.1108
income4	0.1117	<0.001	0.0810	0.1424
welfare	-0.0373	<0.001	-0.0477	-0.0269

Table A10: Marginal Effects for Mild Risk Conditional Probability

Conditional probability of Mild Risk of Nicotine Dependence given selection

Marginal effects are *average* marginal effects, evaluated at observed covariates for each respondent

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
female	-0.01934	<0.001	-0.02344	-0.01524
black	-0.02247	<0.001	-0.02688	-0.01806
hispanic	-0.04708	<0.001	-0.05311	-0.04104
age	0.00344	<0.001	0.00281	0.00407
age2	-0.00006	<0.001	-0.00007	-0.00005
separated	0.02990	<0.001	0.02490	0.03490
midwest	0.00542	0.03611	0.00035	0.01050
south	-0.00271	0.25617	-0.00740	0.00197
west	-0.01182	<0.001	-0.01714	-0.00651
high_school	-0.01537	<0.001	-0.02051	-0.01023
some_college	-0.02564	<0.001	-0.03166	-0.01963
college	-0.04435	<0.001	-0.05158	-0.03713
graduate	-0.07159	<0.001	-0.08221	-0.06097
income1	-0.05209	<0.001	-0.06912	-0.03505
income2	-0.04954	<0.001	-0.06659	-0.03250
income4	-0.06128	<0.001	-0.07886	-0.04371
welfare	0.01187	<0.001	0.00699	0.01675

Table A11: Marginal Effects for Moderate Risk Conditional Probability

Conditional probability of Moderate Risk of Nicotine Dependence given selection

Marginal effects are *average* marginal effects, evaluated at observed covariates for each respondent

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
female	-0.02166	<0.001	-0.02450	-0.01881
black	-0.02415	<0.001	-0.02723	-0.02107
hispanic	-0.04101	<0.001	-0.04464	-0.03738
age	0.00465	<0.001	0.00422	0.00508
age2	-0.00006	<0.001	-0.00006	-0.00005
separated	0.02158	<0.001	0.01840	0.02477
midwest	0.00335	0.06228	-0.00017	0.00687
south	-0.00369	0.02186	-0.00684	-0.00054
west	-0.01242	<0.001	-0.01601	-0.00883
high_school	-0.00964	<0.001	-0.01311	-0.00618
some_college	-0.01589	<0.001	-0.01979	-0.01199
college	-0.03686	<0.001	-0.04117	-0.03255
graduate	-0.06353	<0.001	-0.06926	-0.05779
income1	-0.01488	0.00394	-0.02500	-0.00476
income2	-0.01203	0.01695	-0.02190	-0.00215
income4	-0.02273	<0.001	-0.03274	-0.01272
welfare	0.01462	<0.001	0.01114	0.01810

Table A12: Marginal Effects for Severe Risk Conditional Probability

Conditional probability of Severe Risk of Nicotine Dependence given selection

Marginal effects are *average* marginal effects, evaluated at observed covariates for each respondent

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
female	-0.01644	<0.001	-0.01884	-0.01403
black	-0.01856	<0.001	-0.02115	-0.01597
hispanic	-0.03379	<0.001	-0.03693	-0.03065
age	0.00335	<0.001	0.00295	0.00375
age2	-0.00005	<0.001	-0.00005	-0.00004
separated	0.01909	<0.001	0.01648	0.02169
midwest	0.00316	0.03178	0.00028	0.00605
south	-0.00265	0.04709	-0.00527	-0.00003
west	-0.00961	<0.001	-0.01260	-0.00663
high_school	-0.00904	<0.001	-0.01184	-0.00623
some_college	-0.01498	<0.001	-0.01810	-0.01186
college	-0.03089	<0.001	-0.03444	-0.02734
graduate	-0.05200	<0.001	-0.05697	-0.04703
income1	-0.02117	<0.001	-0.02873	-0.01361
income2	-0.01900	<0.001	-0.02651	-0.01150
income4	-0.02768	<0.001	-0.03537	-0.02000
welfare	0.01080	<0.001	0.00792	0.01367