Received: 24 February 2017 | Revised: 8 September 2017 | Accepted: 13 September 2017

DOI: 10.1002/hec.3614

SUPPLEMENT PAPER

WILEY Health Economics

Behavioral responses to surveys about nicotine dependence

Glenn W. Harrison

Center for the Economic Analysis of Risk, Department of Risk Management & Insurance, Robinson College of Business, Georgia State University, Atlanta, GA, USA

Correspondence
Glenn W. Harrison, Center for the Economic Analysis of Risk, Department of Risk Management & Insurance, Robinson College of Business, Georgia State University, P.O. Box 4036, Atlanta, GA 30302-4036, USA.
Email: gharrison@gsu.edu

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.

KEYWORDS

nicotine dependence, population prevalence, sample selection bias, survey methods

1 | INTRODUCTION

The population prevalence of nicotine dependence (ND) 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 Diagnostic and Statistical Manual of Mental Disorders (*DSM-IV*) and *DSM-5* (American Psychiatric Association, 1994, 2013) clinical criteria for 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 that 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 ND over time.

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.

[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.

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 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.

2 | METHODS

2.1 | Data

There exist many survey instruments designed to measure ND, such as the Fagerström Test for ND (e.g., Heatherton, Kozlowski, Frecker, & Fagerström, 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 2000/2001, 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 three or more of seven 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.

2.2 | Statistical procedure

The sample selection models developed by Heckman (1976, 1979) provide one statistical 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.

[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 five times in any 1 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 do not know well." The survey considered here used face-to-face interviews.

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 nonaddiction that a binary classification requires, and is consistent with increased diagnostic emphases on "severity" in *DSM-5*.

For ND, respondents can be classified into four categories: **No Risk** individuals report no *DSM-IV* criteria or not being asked about them; **At Risk** individuals report one or two *DSM-IV* criteria; **High Risk** individuals report three or four *DSM-IV* criteria; and **Severe Risk** individuals report five or more *DSM-IV* criteria.[4] 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 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.[5]

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

$$Y_j^* = \beta_j^{\top} X_j + U_j \qquad j = 1, 2 \tag{1}$$

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

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

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

[4]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.

[5]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) 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.

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 ND. 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 ND.[6]

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, because 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 defense[7] of the use of sample selection corrections to estimates of HIV prevalence in Bärnighausen, Bor, Wandira-Kazibwe, and Canning (2011a) came from Bärnighausen, Bor, Wandira-Kazibwe, and Canning (2011b) on the grounds that they had access to ideal exclusionary 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. 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 ND. 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.[8] An online appendix describes all covariates used, along with summary descriptive statistics.

[6]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 $R \times R$ 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.

[7]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).

[8]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.

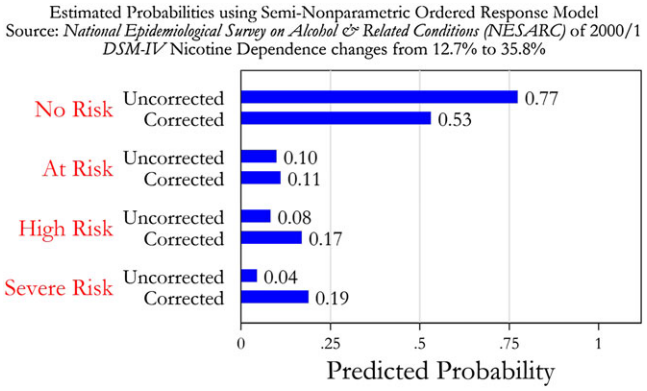

FIGURE 1 Predicted prevalence of nicotine dependence for current daily smokers, with and without sample selection correction. DSM-IV = Diagnostic and Statistical Manual of Mental Disorders [Colour figure can be viewed at wileyonlinelibrary.com]

3 | RESULTS

Figure 1 reports estimates of population prevalence for ND that ignore sample selection and estimates that correct for it.[9] 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%.[10] The 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.

Figure 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. Because 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 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

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

[10]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 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 De Luca 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.

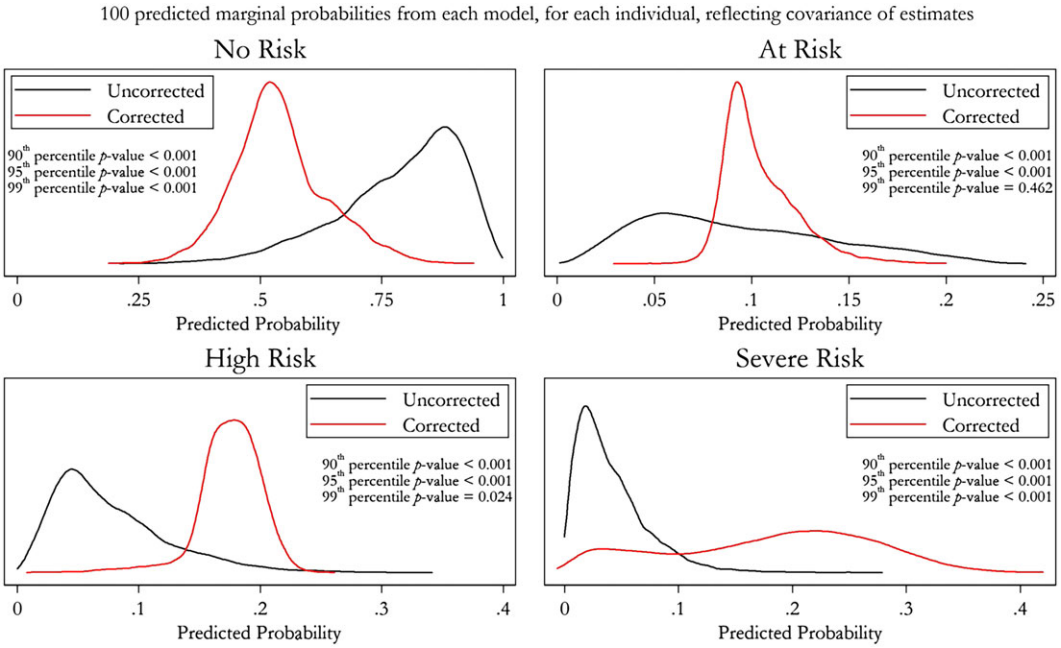

FIGURE 2 Statistical significance of sample selection corrections for nicotine dependence [Colour figure can be viewed at wileyonlinelibrary.com]

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 .001 or lower.[11] There are statistically significant increases in the High Risk and Severe Risk categories. For High Risk, the 95th percentile of *p*-values is .002 or lower, and for Severe Risk the 99th percentile of *p*-values is also .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.[12] It is natural to start by looking at the onset of daily smoking, because 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 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.

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

[12]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.

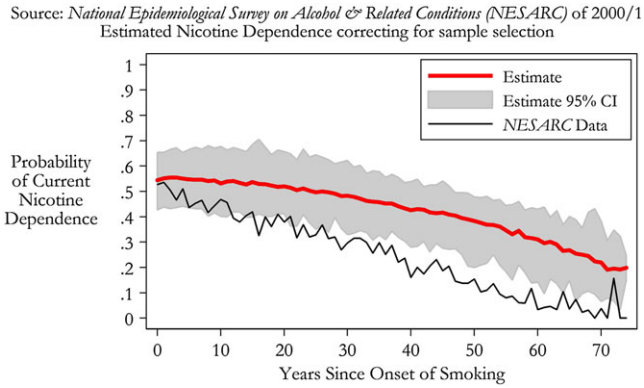

FIGURE 3 Probability of current nicotine dependence by years since onset of smoking for current daily smokers [Colour figure can be viewed at wileyonlinelibrary.com]

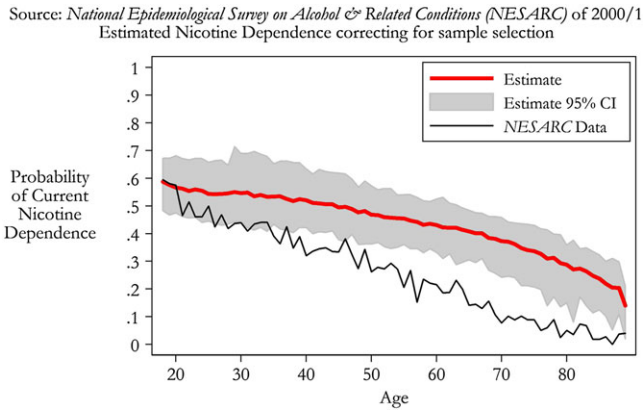

FIGURE 4 Probability of current nicotine dependence by age of current daily smokers [Colour figure can be viewed at wileyonlinelibrary.com]

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, & Rutström, 2009, Harrison, Lau, & Yoo, 2014).[13]

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 ND is the vast literature comparing self-reports of smoking with objective measures that are known to be correlated with exposure to smoke (see Gorber,

[13]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 comorbidities between psychiatric disorders.

Schofield-Hurwitz, Hardt, Levasseur, & Tremblay, 2009 for a survey). An excellent example comes from the Canadian Health Measures Survey. 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 1 day and 6 months after the initial survey responses. Using data from the 2007–2009 wave of the Canadian Health Measures Survey, Wong et al. (2012) show consistency between self-reports and objective measures, using a cutoff in cotinine of 50 ng/ml to define "current smokers," which includes current daily smokers and 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.[14] Information of this kind could be used to generate statistical calibration functions to better infer smoking prevalence[15] and to identify which characteristics are associated with reporting bias.[16]

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, because 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 ND 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, & 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 four largest manufacturers of cigarettes in the United States. The MSA (Section IV) led to the public release of documents uncovered during the discovery process of the litigation, resulting in over 14 million documents being collated 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 (Section 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.[17] 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

[14]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 nonsmokers who did not met the objective threshold for being declared a smoker.

[15]Gorber, Shields, Tremblay, and McDowell (2008) demonstrate this possibility for self-reports of obesity, using data from the CHMS.

[16]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.

[17]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.

were parties to the MSA have no legal recourse 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.

ACKNOWLEDGEMENTS

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.

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. Harrison is also affiliated with the School of Economics of the University of Cape Town.

ORCID

Glenn W. Harrison http://orcid.org/0000-0003-1837-8020

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders IV (DSM-IV)*. Washington DC: APA Press.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders 5 (DSM 5)*. Washington DC: APA Press.

Bärnighausen, T., Bor, J., Wandira-Kazibwe, S., & Canning, D. (2011a). Correcting HIV prevalence estimates for survey nonparticipation using Heckman-type selection models. *Epidemiology*, *22*(1), January, 27–35.

Bärnighausen, T., Bor, J., Wandira-Kazibwe, S., & Canning, D. (2011b). Interviewer identity as exclusion restriction in epidemiology. *Epidemiology*, *22*(3), May, 446.

Cawley, J., & Choi, A. (June 2015) Health disparities across education: The role of differential reporting error, *IZA Discussion Paper 9141*, Institute for the Study of Labor, Bonn.

Chaix, B., Bilaudeau, N., Thomas, F., Havard, S., Evans, D., Kestens, Y., & Bean, K. (2011). Neighborhood effects on health: Correcting bias from neighborhood effects on participation. *Epidemiology*, *22*(1), January, 18–26.

Coller, M., Harrison, G. W., & McInnes, M. M. (2002). Evaluating the tobacco settlement: Are the damages awards too much or not enough? *American Journal of Public Health*, *92*(6), 984–989.

De Luca, G. (2008). SNP and SML estimation of univeriate and bivariate binary-choice models. *Stata Journal*, *8*(2), 190–220.

De Luca, G., & Perotti, V. (2011). Estimation of ordered response models with sample selection. *Stata Journal*, *11*(2), 213–239.

Derthick, M. A. (2011). *Up in smoke: From legislation to litigation in tobacco politics* (Third ed.). Washington, DC: CQ Press.

Gallant, A. R., & Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica*, *55*(2), March, 363–390.

Geneletti, S., Mason, A., & Best, N. (2011). Commentary: Adjusting for selection effects in epidemiologic studies: Why sensitivity analysis is the only solution. *Epidemiology*, *22*(1), January, 36–39.

Gorber, S. C., Schofield-Hurwitz, S., Hardt, J., Levasseur, G., & Tremblay, M. (2009). The accuracy of self-reported smoking: A systematic review of the relationship between self-reported and cotinine-assessed smoking status. *Nicotine & Tobacco Research*, *11*, 12–24.

Gorber, S. C., Shields, M., Tremblay, M. S., & McDowell, I. (2008). Feasibility of establishing correction factors to adjust self-reported estimates of obesity. *Health Reports*, *19*(3), September, 71–82. Statistics Canada Catalogue #82-003-XPE

Grant, B. F., Dawson, D. A., Stinson, F. S., Chou, S. P., Kay, W., & Pickering, R. (2003). 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*, 7–16.

Grant, B. F., Hasin, D. S., Chou, S. P., Stinson, F. S., & Dawson, D. A. (2004a). 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, 1107–1115.

Grant, B. F., Stinson, F. S., Dawson, D. A., Chou, S. P., Ruan, W. J., & Pickering, R. P. (2004b). 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, 361–368.

Harrison, G. W., Lau, M., & Yoo, H. I. (2014). 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.

Harrison, G. W., Jessen, L. J., Lau, M., & Ross, D. (2017). Disordered gambling prevalence: Methodological innovations in a general Danish population survey. *Journal of Gambling Studies*, https://doi.org/10.1007/s10899-017-9707-1.

Harrison, G. W., Lau, M. I., & Rutström, E. E. (2009). Risk attitudes, randomization to treatment, and self-selection into experiments. *Journal of Economic Behavior and Organization*, *70*(3), June, 498–507.

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerström, K.-O. (1991). The Fagerström test for nicotine dependence: A revision of the Fagerström tolerance questionnaire. *British Journal of Addiction*, *86*, 1119–1127.

Heckman, J. J. (1976). 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*, *5*, 475–492.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–162.

Hernán, M. A., Hernández-Diaz, S., & Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology*, *15*(5), September, 615–625.

Narrow, W. E., Rae, D. S., Robins, L. N., & Reiger, D. A. (2002). Revised prevalence estimates of mental disorders in the United States: Using a clinical significance criterion to reconcile 2 surverys' estimates. *Archives of General Psychiatry*, *59*, 115–123.

Rothman, K. J., Greenland, S., & Lash, T. L. (2012). *Modern epidemiology* (3rd ed.). New York: Lippincott, Williams & Wilkin.

Stewart, M. B. (2004). Semi-nonparametric estimation of extended ordered probit models. *Stata Journal*, *4*(1), 27–39.

Stewart, M. B. (2005). A comparison of semiparametric estimators for the ordered response model. *Computational Statistics & Data Analysis*, *49*, 555–573.

Wong, S. L., Shields, M., Leatherdale, S., Malaison, E., & Hammond, D. (2012). Assessment of validity of self-reported smoking status. *Health Report*, *23*(1), March, 1–7. Statistics Canada Catalogue #82-003-XPE.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Harrison GW. Behavioral responses to surveys about nicotine dependence. *Health Economics*. 2017;26(S3):114–123. https://doi.org/10.1002/hec.3614