The Risk of Gambling Problems

in the General Population: A Reconsideration

by

Glenn W. Harrison, Morten Lau and Don Ross†

July 2018

ABSTRACT.

Gambling behavior is pervasive, apparently growing, and of methodological and substantive interest to economists. We examine the manner in which the population prevalence of disordered gambling has been estimated. General population surveys have deepened our knowledge of the population prevalence of gambling disorders, as well as the manner in which gambling disorder is associated with other mental health problems. However, we identify a fundamental bias in the manner in which these surveys have been used to draw inferences about the general population prevalence of gambling problems, due to a behavioral response to seemingly innocuous “trigger,” “gateway” or “diagnostic stem” questions in the design of surveys. Formal modeling of the latent sample selection behavior generated by these trigger questions leads to dramatically different inferences about population prevalence and comorbidities with other psychiatric disorders. The population prevalence of problem or pathological gambling in the United States is inferred to be 7.7% rather than 1.3% when this behavioral response is ignored. Comorbidities are inferred to be much smaller than the received wisdom, particularly when considering the marginal association with other mental health problems rather than the total association. The issues identified here apply, in principle, to every psychiatric disorder covered by these surveys, and not just gambling disorder. We discuss ways in which these behavioral biases can be mitigated in future surveys.

† Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Harrison); Copenhagen Business School, Denmark (Lau); and School of Philosophy and Sociology, University College Cork, Ireland; School of Economics, University of Cape Town, South Africa; and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Ross). Harrison is affiliated with the School of Economics, University of Cape Town, and Lau is affiliated with Durham University. E-mail contacts: gharrison@gsu.edu, mla.eco@cbs.dk and don.ross@uct.ac.za. We are grateful to the U.S. National Institute on Alcohol Abuse and Alcoholism, the British Gambling Commission, the Victorian Responsible Gambling Foundation, the Victorian Department of Justice and Regulation and Statistics Canada for providing access to survey data, and to the Danish Social Science Research Council (Project #12-130950) for financial support.
Table of Contents

1. Estimates for the United States from NESARC ........................................... -8-
   A. Comorbidities .................................................................................. -8-
   B. Sample Selection ............................................................................. -10-
   C. The Hierarchy of Gambling Disorders ............................................. -15-

2. Sample Selection Bias and Gambling Survey Screens ................................ -19-
   A. The Evolution of Trigger Questions ................................................. -19-
   B. Mitigating the Effects of Trigger Questions ..................................... -24-

3. Conclusions ........................................................................................... -28-

References .................................................................................................. -42-

Appendix A: Classifying Pathological Gamblers in NESARC (NOT FOR PUBLICATION) . . . -A1-
Appendix B: Additional Documentation of Results (NOT FOR PUBLICATION) ........ -A10-
Appendix C: Formal Statistical Model (NOT FOR PUBLICATION) ..................... -A18-
Gambling behavior is pervasive across human history and cultures, is increasingly widespread after a wave of regulatory liberalization in the last 20th century, and is of methodological and substantive interest to economists. We question the measurement of the prevalence of gambling problems in the general population, concluding that it appears to be significantly larger than claimed.

In the United States, commercial casino revenue has steadily increased, from $27.1 billion in 2001 to $40.2 billion in 2015 (Shwartz [2016]). State and local lottery revenue has increased from $770 million in 1977 to $23.7 billion in 2013 (Tax Policy Center [2016]). And sports betting in states was deemed legal in a landmark ruling in May 2018 by the Supreme Court of the United States.¹ Scholarly research consistently finds high shares of commercial gambling revenue to be derived from proportions of populations that are much smaller that the large proportion who occasionally or frequently gamble. For the United States 15% of revenue derives from 0.5% of the population, for Canada 23% derives from 4.2% of the population, for Australia 33% derives from 2.1% of the population, and for New Zealand 19% derives from just 1.3% of the population.² It is primarily among the ranks of these high-spending gamblers that one finds those who currently have, or are at greatest risk for having, clinically diagnosable gambling problems.³ And the largest share of casino gambling floor revenue now derives from electronic slot and poker machines, which are strikingly characterized as constituting an Addiction by Design by Schüll [2012].

The popularity of gambling, given the near certainty of monetary loss that attends any

¹ In Murphy v. NCAA the Supreme Court struck down a 1992 Federal ban on sports betting. Soon after, Nevada, Delaware and New Jersey legalized sports betting, with several states currently in various stages of legislative approval. The 1992 ban “grandfathered in” sports betting in casinos in Las Vegas, and limited sports lotteries and sports pools in Delaware, Montana and Oregon.

² See Goldstein et al. [1999], Williams and Wood [2004], Australian Productivity Commission [1999] and Abbott and Volberg [2000], respectively. Definitions of those most at risk of gambling-control problems vary across studies, for reasons we discuss in detail, but all are conventional in the extant literature.

³ Of course, welfare losses from gambling are not confined to the clinically diagnosable. But given that gambling also produces welfare gains in the form of hard-to-measure entertainment value, multipliers on corporate profits, tax revenues, and offset contributions to public infrastructure in exchange for licenses, estimation of net costs and benefits is a complex exercise. Walker [2013] provides a recent attempt.
significant level of indulgence in it by a consumer, has long presented a puzzle for economists, and its classic rationalization by Friedman and Savage [1948] is among the core models in the economics of risk. Addictive gambling, like addiction more generally, presents a particularly acute methodological challenge to economists who rely on axioms that posit consistent preference, because most gambling addicts simultaneously spend money and also expend resources trying to gamble less (Ross et al. [2008]). There is still no consensus solution to this challenge, of a kind that could turn welfare assessments of gambling regulation and public health interventions aimed at reducing gambling addiction into mainly technical, as opposed to partly philosophical, exercises. Clearly, one crucial prerequisite for building a successful model for such welfare estimations is accurate estimation of the prevalence of current gambling problems, and of the frequency and determinants of vulnerability to such problems. There have been numerous surveys of disordered gambling prevalence conducted around the world, but these have not generally benefitted from participation in their design or interpretation by economists.

We step back and examine in detail the manner in which the population prevalence of disordered gambling has been estimated by psychologists and psychiatric researchers. Surveys of disordered gambling have traditionally used screens designed to detect individuals who engage in gambling activity that might lead them to clinically “present” and meet criteria for diagnosis of a psychiatric disorder. This is a valid scientific goal for the design, calibration and application of such surveys, although it is not the only possible goal or the most interesting for broader public health assessments.4 We reconsider the manner in which inferences about gambling problems in the general population are made based on these surveys. We suggest that there are different kinds of inferences possible than have traditionally been emphasized, and that there is a recurring, major sample selection

4 Kessler and Pennell [2015; p. 144ff.] provide a valuable review of the historical evolution of survey research on mental disorders.
bias that has not been accounted for. When that bias is corrected we infer significantly greater prevalence of gambling disorders, and notably fewer comorbidities with other mental health problems. Thus we contribute to isolating gambling disorder as a partly discrete public health problem to which policies can be specifically targeted and their efficiency evaluated.

Most of the inferences that have been drawn based on analysis of these surveys have concerned general population prevalence, socio-demographic correlates, and comorbidities. They have typically focused on the binary classification of individuals as “disordered,” “pathological” or “problem” gamblers, or not, where these terms are defined either directly or approximately in terms of DSM-IV (American Psychiatric Association [1994]) clinical criteria.5

The term “pathological gambler” was introduced in DSM-III (American Psychiatric Association [1980]). There, and in DSM-IV, Pathological Gambling was classified among “Impulse Control Disorders Not Elsewhere Classified.” Other disorders in this category were Compulsive Hair Pulling (Trichotillomania), Intermittent Explosive Disorder, Kleptomania, and Pyromania. By the time of DSM 5 (American Psychiatric Association [2013]), growing evidence of behavioral, etiological, and clinical response similarities between pathological gambling and addictive substance dependencies had been documented in the scientific literature; Ross et al. [2008; Chapter 5] survey this evidence. Consequently the disorder that had been referred to as pathological gambling was reclassified in DSM 5 among “Substance-Related and Addictive Disorders.” All other disorders in that category are dependencies on addictive substances. At the same time, pathological gambling was re-named “Gambling Disorder.”

The standard scoring rule for classification was not amended: pre-2013 “Pathological Gamblers” and current “Disordered Gamblers” are people who report 5 or more of 10 DSM criteria

5 The DSM is the Diagnostic and Statistical Manual of Mental Disorders, published by the American Psychiatric Association.
over a reference frame of the past 12 months. Thus the clinical conception of the disorder has not changed, notwithstanding revision in the underlying theory of its etiology and relationship to other mental disorders. The new terminology was partly motivated by this reconceptualization, and partly by historical inconsistency in use of the term “pathological gambling” as sometimes contrasted with, and sometimes used as synonymous with, “problem gambling” (confusion flagged as unfortunate by the Committee on the Social and Economic Impact of Pathological Gambling [1999]).

A number of researchers, such as Petry et al. [2014; p.494], also argued that “pathological gambling” should be retired from professional use because it is associated with social stigmatization. However, the implication of this motivation, that “gambling disorder” should replace “pathological” gambling, and perhaps also “problem gambling,” has not become consensus practice among researchers or clinicians. Some recent research, following a sub-tradition of using “problem gambling” to refer to the presence of “pathological gambling” at pre-clinical levels, and as denoting an early or risk-indicating possible precursor stage to full-blown “pathological gambling,” have explicitly applied “pathological gambling” to the most severe manifestation of “gambling disorder” and applied “problem gambling” to less acute cases of “gambling disorder” (Pietrzak et al. [2007], Algeria et al. [2009] and Nower et al. [2013]). These researchers have applied a scoring rule to the DSM criteria due to Fisher [2000], according to which a sub-clinical “problem gambler” is someone who records between 1 and 4 positive responses with at least one of these being to one of the final 3 DSM-IV criteria, thus leaving all three phrases in play, and using “gambling disorder” to denote a larger class of people than the DSM 5 does. Finally, the most commonly used measurement screen at present, the Problem Gambling Severity Index (PGSI) of Ferris and Wynne [2001], uses the term “problem gambling” to group together all gamblers within the highest severity category; in effect the PGSI uses

---

6 These are: (8) Have you been forced to go beyond what is strictly legal, in order to finance gambling or to pay gambling debts?; (9) Have you risked or lost a significant relationship, job, educational or career opportunity because of gambling?; (10) Have you sought help from others to provide money to relieve a desperate financial situation caused by gambling?
“problem gambling” as a synonym for the category “gambling disorder” in DSM 5. Thus PGSI-designated “problem gamblers” and “problem gamblers” in the usage of Pietrzak et al. [2007], Algeria et al. [2009] and Nower et al. [2013] are, at least in intention, disjoint sets!

In light of this semantic chaos, researchers must stipulate terminological policies. Henceforth, where we refer to the clinical phenomenon ex cathedra we will follow DSM 5 and use “gambling disorder” (GD). We will refer to a representative person who has acquired the condition as a “disordered gambler” (DG). Where we refer to previous work set in clinical contexts that used either “pathological” or “problem” gambling without distinguishing them, or intending that they be distinguished (for example, in work applying the PGSI), we will anachronistically use the terms “gambling disorder” and “disordered gambler.” Where we are referring to a context in which “problem gambling” and “pathological gambling” are distinguished, with the former denoting a pre-clinical or warning state for the latter, we retain the distinction and use these older terms. Thus, in our usage, “pathological gambler” will always be intended to pick out the people classified as “disordered gamblers” by DSM-IV and DSM 5. Finally, when we talk about harmful consequences of gambling outside the clinical context we use “gambling problems” as a non-technical term of everyday English.

An original goal of DSM 5 was to shift focus away from categorical classifications emphasized in DSM–III and DSM–IV (e.g., “pathological/non-pathological”) to continuous measures, understood as probing continua between normal and disordered functioning. However, the American Psychiatric Association ultimately decided to defer this ambition, and GD continues to be clinically regarded as a pathology from which a person either suffers or does not. We implicitly consider that classification, but expand the analysis to include the range of gambling problems as an ordered hierarchy. Our interest is in the latent continuum of gambling problems, as a complement to studying binary classifications.
with thresholds.\textsuperscript{7} This interest corresponds to our ultimate focus, as economists concerned with the general impact on welfare of gambling and public health policy, on evaluating the severity of all problems associated with gambling, which include but are not limited to the form of addiction that \textit{DSM 5} labels as GD.\textsuperscript{8}

In Section 1 we reconsider inferences from the \textit{National Epidemiologic Survey on Alcohol and Related Conditions} (NESARC) in the United States. The first wave of NESARC was conducted in 2000 and 2001, and had a sample of 43,093 individuals.\textsuperscript{9} The instrument for measuring gambling problems was based on the \textit{DSM-IV} criteria, and \textit{DSM-IV} criteria were likewise used for the instruments measuring other major psychiatric disorders.\textsuperscript{10}

The most significant statistical issues arise from the difficulty of drawing inferences about GD prevalence and comorbidity when one attempts to account for the sample selection bias of “trigger,” “gateway” or “diagnostic stem” questions. Such questions ask whether a respondent has ever gambled more frequently than some threshold rate or number of occasions, and/or whether they have ever gambled away more than some threshold amount of money on any single occasion. Only respondents who report meeting the relevant thresholds are asked the remaining gambling screen questions. A

\footnote{Examples of the many studies of the continuum of gambling disorders include Toce-Gerstein, Gerstein and Volberg [2003] and Blanco, Hasin, Petry, Stinson and Grant [2006].}

\footnote{Harrison and Ng [2016] is an example of our general approach, applied to the problems of making decisions over an insurance product to evaluate the welfare cost to the individual of observed choices. That cost is measured by the foregone income-equivalent of the observed choice compared to what a latent structural theory predicts that the individual should have made. Calculating this income cost, which in the case of insurance arguably maps relatively straightforwardly onto welfare costs, requires a different set of data about the individual than one finds in surveys, but the end result is more usefully compared to non-binary measures of the severity of behavior.}

\footnote{The second wave of the NESARC was conducted in 2004/5, and was a longitudinal panel of 34,653 re-interviews from the first wave. The third wave was conducted in 2012/13, with a fresh sample of 36,309 individuals. Gambling prevalence questions were removed from waves 2 and 3 of the NESARC. Our 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.}

\footnote{A comparable national survey that could also be evaluated in the same manner is the National \textit{Comorbidity Survey Replication} conducted in the United States between 2001 and 2003 with a primary sample of 9,282 individuals. We discuss the \textit{Canadian Community Health Survey of Mental Health and Well-Being} of 2002 and the \textit{British Gambling Prevalence Survey} of 2010 below.}
main motivation for use of trigger questions is not to irritate respondents by asking them about gambling problems after they have effectively said that they are not regular gamblers, or perhaps not gamblers at all. This motivation is particularly easy to appreciate in the case of surveys such as the NESARC, which address multiple potential disorders; the surveyor does not want to risk reduced cooperation on other survey modules by annoying respondents about gambling problems they (apparently) manifestly do not have.

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.\(^\text{11}\) 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 \textit{a priori} reason to expect an effect on inferences about gambling risks. On the other hand, if someone wants to hide their gambling problems, they might reasonably choose to lie in response to the trigger question. Indeed, 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 GD or should be classified as a DG! There are

\(^{11}\) Hernán, Hernández-Díaz 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]. To our knowledge, there are no applications of epidemiological corrections for these biases to general population surveys with trigger questions, although there are recognitions of their potential importance (Tam, Midanik, Greenfield and Caetano [1996] and Tam and Midanik [2000]). Caetano [2001; p.1543] editorialized on this issue in a clear fashion: “So, are survey respondents different from non-respondents in their use of alcohol and illicit drugs? The answer from the small number of studies mentioned above seems to be positive. But are my critics right in assuming that non-respondents are more likely than respondents to be drinkers, heavier drinkers or dependent on alcohol and illicit drugs? The evidence then suggests that, to use a common American expression, the ‘jury is still out.’ This is so partly because for the past 40 years those of us facing the critics have been complacent about the validity of survey research. The uncertainty regarding selective non-response should not justify the apparent lack of attention to the issue.”
no perfect statistical methods to correct for this bias, but the bias appears to be significant in the case of several major, influential surveys of gambling problems that used trigger questions. We therefore take some time in section 2 to review the rationale for these trigger questions, and note with some surprise the “aggressive” rhetoric sometimes used to defend them. We suspect that the strength of these defenses is thought to be justified by an expectation that they have no effect on inference, and the efficiency gains in the time needed to conduct surveys that are apparent from their use. Section 3 draws some conclusions, including recommendations for future survey design and analysis.

1. Estimates for the United States from NESARC

A. Comorbidities

The prevailing view is that GD typically co-occurs with a variety of other mental disorders. Petry et al. [2005; p.564] evaluated this using NESARC data and concluded that GD is “highly comorbid with substance use, mood, anxiety, and personality disorders, suggesting that treatment for one condition should involve assessment and possible concomitant treatment for comorbid conditions.” Panel A of Table 1 replicates their methods and essentially obtains the same results, using a logistic specification. All calculations with the NESARC correct for the complex sampling design.

---

12 Their analysis, and most of those using the NESARC to study DG, suffers from an unfortunate coding error explained in Appendix A. There are in fact 207 respondents that meet the DSM-IV criteria, not the 195 used in most studies. The incorrectly coded classification had 21 respondents that should have been classified as DGs, and 9 that should not have been so classified. The effect is to change estimates slightly. We only use the correct DSM-IV classification of pathological gambling from the NESARC. None of our qualitative conclusions are affected by using the incorrect classification.

13 The NESARC used a three-stage sampling design, with a sampling frame of adults aged 18 and over in non-institutionalized settings. Stage 1 was primary sampling unit (PSU) selection using the PSUs from the Census 2000/2001 Supplementary Survey, a national survey of 78,300 households per month. Stage 2 was household selection from the sampled PSUs. Finally, in stage 3, one sample person was selected at random from each household. In stage 1 there were 401 PSUs that were so large that they were designated “self representing,” meaning that they were selected with certainty; another 254 PSUs were selected in proportion to 1996 population estimates for each of 9 strata within a state (so there are 10 strata, including the state). Self-representing PSUs within a state are correctly treated as being selected with certainty, and hence contributing nothing to the estimated standard error as a PSU.
In each row the independent binary variable is whether the respondent is defined as having the indicated psychiatric disorder or not. Petry et al. [2005] examine the risk of being what we would now call a DG, and this is the sole risk level used in Table 1. Each of the odds ratio (OR) estimates in Panel A are much greater than 1, and statistically significantly greater than 1: the lower bound of the 95% confidence interval is well above 1.

These analyses of comorbidities examine “total effects” rather than “marginal effects.” We say that one has measured the total effect of some secondary psychiatric disorder X on the focus disorder Y when there are no controls for the presence of other psychiatric disorders A, B, C … etc. The marginal effect of psychiatric disorder X is measured when one controls for the presence of other psychiatric disorders. Both types of effects can be of interest for public health and clinical purposes, and answer different questions. The total effect answers a question along these lines: “If all I know about a group of people is that they abuse alcohol, how likely is it that they are also DGs?” Another total effect question might be, “If all I know about a group of people is that they are chronically depressed, how likely is it that they are also DGs?” Assume, as is the case, that both total effects are positive and statistically significant. The marginal effect answers a different question, of the following kind: “If I know that people abuse alcohol and/or are chronically depressed, what is the incremental correlation of each disorder with their also suffering from GD?” It could be that the incremental correlation of alcohol abuse is low or non-existent and the incremental correlation of chronic depression is high. These particular correlations suggest, but of course do not prove, that there is causality from chronic depression to GD, none from alcohol abuse to GD, and some from chronic depression to alcohol abuse (or vice versa). If this suggestion is correct, it has direct implications for

---

14 A constant term is always employed as well. This is what Petry et al. [2005] refer to as “model 1,” where there are no additional covariates added.

15 Epidemiologists often report “adjusted odds ratios,” which control for covariates. Typically the list of covariates is very small.

16 We know well the dangers of inferring causality from correlations, and indeed this concern is why many modern surveys of mental health take time to ask additional questions about “age of onset.” This
treatment for GD. We would argue that marginal effects are closer to what we want to learn about from evaluation of general population surveys, at least for purposes of designing and choosing public health interventions, than total effects.

Panel B of Table 1 shows the estimates of comorbidities, focusing on marginal effects and the implied OR. We use the same econometric specification as Panel A, for comparability. The point estimates are much closer to 1 than the total effects, as are the lower bounds of the 95% confidence intervals. In one case, the comorbidity of anxiety and GD, the OR is not statistically significantly different from 1. The upper bound of the 95% confidence interval of marginal effects in Panel B are all well below the lower bound of the 95% confidence interval of total effects in Panel A.

Panels C and D of Table 1 show comparable estimates of total and marginal effects if one includes a long list of socio-economic and socio-demographic covariates. There is a slight lowering of most of the OR compared to Panels A and B, respectively, but no significant change from the conclusions drawn from Panels A and B.

B. Sample Selection

Panel E of Table 1 lists additional covariates from the logistic model estimated to obtain the marginal effects in Panels C and D. To informally motivate the concern with sample selection bias, focus on the OR ratios in Panel E in bold. Imagine we encounter men, Blacks, those separated by divorce or death, people living in the West, those without a college or graduate degree, and those with a personal income over $70k at the time of the survey. The value of these OR estimates, and their

information is particularly important when asking about incidence over lifetime frames, since the correlation might have any one of three temporal sequences (prior, simultaneous, and posterior). This is also why one-shot general population surveys are not the same as clinical evaluations that occur over several meetings, despite the attempt to ask questions about the clinical significance of symptoms. Moreover, it becomes difficult in general surveys to ask enough about the history of an individual to establish if a disorder is “substance-induced,” which is one exclusion criteria used for mood disorders, for example.

17 This is “model 3” of Petry et al. [2005].
statistical significance, tell us that respondents with these characteristics are more likely to be DGs. So suppose we encounter respondents with these characteristics who happened not to respond affirmatively to the trigger question? Without knowing their responses to the trigger question we would be inclined to suspect them of some greater-than-baseline risk of GD, ceteris paribus. The only reason they are not so classified is that their response to the trigger question led to them being assumed to have no current or past gambling problems and, therefore, no risk of GD. This involves two fallacious inferences: first, that no one who says “no” to the trigger question has any current or past gambling problems, and, second, that there are no other potential indicators of risk. We can easily imagine some degree of sample selection bias if the responses to the trigger question are correlated with the characteristics that constitute these additional indicators. This is loose and informal, since it is based on a “chicken and egg” fallacy – we are looking at estimates that ignore this sample selection correction to motivate the possibility of sample selection bias. But as long as we check this with appropriate methods, this motivation is acceptable.

The sample selection models developed by Heckman [1976][1979] meet this need. They require the researcher to specify a sample selection process, characterizing which respondents appear in the main survey and which do not. Typically this is a simple binary matter, so one can specify this process with a probit model. In our case the sample selection consists of some trigger questions we examine in a moment; if the respondent passes these, they are admitted to the main survey and asked the DSM criteria questions. The Heckman approach also requires a model of the data generation process in the main survey. In our case this might consist of a binary choice statistical model.

---

18 This example also points to the logic of the correction for sample selection discussed below. If there is a correlation between the unobserved characteristics that affect one’s selection into the sample and the unobserved characteristics that affect one’s chance of being at risk for gambling problems, then the residuals from equations measuring these two behavioral responses (to the trigger question, and then to the full set of questions) will also be correlated. This correlation of the residuals, or covariance, is used to infer what the responses would have been to the full set of equations if there had not been this systematic selection into the sample responding to the full set of questions. Note that we stress the idea of a “systematic” selection bias, with no presumption that it is a deliberate choice to lie in response to the trigger question.
explaining whether someone meets the DSM threshold for being potentially classified as a DG.

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 specification was Ordinary Least Squares. In our case, at least initially, the main data generating process underlying the classification as a DG is binary, and the same ideas carry over: Van de Ven and Van Praag [1981] is the first application of sample selection to a probit specification of the behavior of interest, and Lee [1983] and Maddala [1983] provide general expositions.

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, the natural first assumption is that the errors are bivariate normal.19 We assume instead a flexible semi-nonparametric (SNP) approach due to Gallant and Nychka [1987], applied to the sample selection model by De Luca and Perotti [2011]. This SNP approach approximates the bivariate density function of the errors by a Hermite polynomial expansion.20

In addition, another 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,

---

19 The methods we use are full maximum likelihood. The “limited information” estimator of Heckman [1976][1979] did not require all of the properties of the bivariate normal distribution. All that was required was that there be a linear relationship between the errors of the two equations, and that the error of the sample selection equation be marginally normal (so that one could calculate the inverse Mills ratio).
20 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. Appendix C presents the formal statistical model.
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\textsuperscript{21} 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
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 \textit{particularly} “problematic” in the absence of \textit{a priori} convincing exclusion restrictions is
that one must rely on having the right econometric specification if the sample selection model is to
correct for sample selection bias. This specification in turn refers to the specification of the two
equations as probit models, and specifically to the assumed bivariate normality of errors.\textsuperscript{22} The
importance of having the right specification of the error distribution also applies even when one does
have exclusion restrictions.

As it happens, there are ways to construct exclusion restrictions in NESARC that have some \textit{a priori}
credibility. For instance, we know the day of the week on which the interview was conducted,
and can condition on Friday, Saturday or Sunday interviews as potentially generating differential
response. We also know how many trigger questions for other disorders a subject had answered

\textsuperscript{21} 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].

\textsuperscript{22} Thus one finds comments such as: “Theoretically, we do not need such identifying variables, but
without them, we depend on functional form to identify the model. It would be difficult for anyone to take
such results seriously because the functional form assumptions have no firm basis in theory.” (StataCorp
[2013; p. 782]). A similar comment from Bärnighausen et al. [2011b; p. 446] in an epidemiological setting is
that the “... performance of a Heckman-type model depends critically on the use of valid exclusion
restrictions...” It is agreed that the functional form assumptions, including the bivariate normal error
assumptions, have no firm basis in theory, but we make such assumptions all the time in other settings. If we
can indeed test them, that would be ideal, but it is not clear why we should in this instance not use them if we
have to. Our view is that these models should be viewed as statistical “canaries in the cave,” in the sense of
pointing to potentially disastrous conditions that warrant immediate investigation. In other words, and to put
the inferential shoe on the other foot, if some estimates show great sensitivity to sample selection corrections
with these assumptions, and some decent effort to find good specifications, then one should not ignore that
evidence because some of the parametric assumptions are untestable.
affirmatively by the time the gambling trigger questions were asked, as one measure of how much time and “patience” had been taken up 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. But such exclusion restrictions do not always arise in other surveys of gambling, even major epidemiological surveys. In general we recommend survey methods that do not require these sorts of tradeoffs (between finding attractive exclusion restrictions and reliance on the assumed stochastic structure for identification), but with existing surveys some tradeoffs are often needed.

To set the stage for the evaluation of sample selection corrections, Table 2 and Figure 1 show the estimated OR between GD and other psychiatric disorders when using the SNP approach rather than the parametric logistic specification. The total comorbidity estimates in Panel A of Table 2 are comparable to those in Panel C of Table 1; similarly, the marginal comorbidity estimates in Panel B of Table 1 are comparable to those in Panel D of Table 1. With the SNP approach, however, the marginal comorbidities are not quite as close to 1 as with the parametric model. However, the same qualitative conclusions about the relationship of total and marginal comorbidity still apply.

Figure 2 shows marginal effects of comorbidities when one undertakes sample selection corrections.23 The covariates used for this exercise are the same full set used in “model 3” of Petry et al. [2005], and are used for both equations.24 In addition, for the sample selection equation, we used a set of 29 variables reflecting recent events in the life of the respondent (e.g., family deaths or illness, job layoff, change in job, problems with neighbors or friends, criminal problems), height and weight, days of the week for the interview, and the number of previous trigger questions answered affirmatively. The variables reflecting life events only referred to the last year or last few months prior to the interview.

23 We undertake sample selection corrections for GD, but not for the other psychiatric conditions. Instead we use the NESARC determinations of diagnosis. An important extension of our approach would be to simultaneously undertake sample selection corrections for all conditions and then assess comorbidity with respect to the corrected diagnoses for all conditions.
24 Appendix B documents these covariates.
to the interview, and we are examining GD incidence across the lifetime frame. Table 3 presents
detailed estimates: for now, focus on Panel C, which shows OR with respect to the GD risk level. The
effect of sample selection corrections is clear: the OR estimates are generally much lower. The
estimated correlation between the two equations in the selection model, a measure of the importance
of sample selection corrections, is -0.19.

C. The Hierarchy of Gambling Disorders

We turn to the hierarchy of gambling disorders, and inferences about general population
prevalence. For example, the PGSI classifies samples into the categories “Non-Gambler,” “Low Risk
for Problem Gambling,” “Moderate Risk for Problem Gambling,” and “Problem Gambler.”

Previous statistical evaluations of these hierarchies have not, to our knowledge, 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 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. We classify respondents into 4 categories: Non-Indicated individuals
have no DSM-IV criteria or were not asked about them; Weakly Indicated individuals meeting 1 or 2
DSM-IV criteria; and Moderately Indicated individuals meeting 3 or 4 DSM-IV criteria.

25 The intended interpretation of risk here is not prospective (the probability of developing GD at
some point in the future). Rather, it is intended as the risk that the respondent would currently be diagnosed
as a DG if he or she participated in a full clinical interview with more reliable discrimination.

26 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
the terminology used in the NESARC, and refer to individuals who meet 5 or more *DSM-IV* criteria as **Pathological Gamblers**.

Figures 3 and 4 report estimates from a SNP ordered response model that ignores sample selection and estimates that correct for it. We use the estimates from these models to predict the fraction of the population in each of our four categories above. As a control, it is useful to note that the fractions of the population from the raw data found in each *DSM-IV* response number “bin” are recovered by the estimated ordered response model when we do not correct for sample selection: 94.6% Non-Indicated, 4.0% Weakly Indicated, 0.9% Moderately Indicated, and 0.4% Pathological Gamblers. Hence we know that the base statistical model we have estimated is not biased relative to the raw data, as we have binned it. These base predictions are referred to as the Uncorrected predictions in Figures 3 and 4. We therefore find a common result, that the prevalence of GD is around 0.4%. To the extent that our “Moderately Indicated” individuals are taken to approximately correspond to what some researchers (e.g., Pietrzak et al. [2007], Algeria et al. [2009] and Nower et al. [2013]) categorize as sub-clinical “Problem Gamblers,” the sum of the two most troubled categories produces a figure of 1.3%, familiar from the GD prevalence literature. The Corrected predictions, allowing for sample selection biases, are again dramatic. The fraction of Weakly Indicated increases from 4.0% to 8.3%, the fraction of Moderately Indicated increases from 0.9% to 3.9%, and the fraction of Pathological Gamblers increases from 0.4% to 3.8%. Hence prevalence of Pathological Gamblers plus Moderately Indicated is 7.7% when sample selection bias is corrected, compared to

---

stressed by Narrow, Rae, Robins and Reiger [2002]). There are no such criteria for GD evaluation in *DSM 5* since the symptoms themselves are viewed as evidence of “clinically significant impairment or distress” (American Psychiatric Association [2013; p. 585]). However, *DSM-III, DSM-IV* and *DSM 5* all contain exceptions for anyone whose gambling behavior is not “better explained” by a manic episode. This exclusion criteria is also only asked in surveys if someone met the threshold level of symptoms. For NESARC there are only 25 (7) out of 68 respondents to this question that said that any (all) of the times they gambled happened “during a period when they felt extremely excited, extremely irritable or easily annoyed.” These respondents constitute only 0.042 (0.016) of a percentage point of the population. For consistency of interpretation across the hierarchy, we do not apply this exception.
1.3% when no correction is applied.

It is worth stressing that this result is not simply because the sample selection model predicts that more people will get through the gateway of the trigger question, although it does predict that. The observed fraction being selected by their responses to that question is 27%, 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 58%. 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 4 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 4, for the “Non-Indicated” category of gambling risk. The Uncorrected distribution of predictions reflects the results of simulating 100 random draws for each NESARC respondent from the predicted marginal probability of Non-Indicated, using the estimated

---

27 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 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.

28 The survey of gambling disorders in the Canadian Community Health Survey (CCHS) of Mental Health and Well-Being of 2002 illustrates this point perfectly. Their gateway questions resulted in only 1,754 of 36,884 subjects being asked the full set of questions from the Canadian Problem Gambling Index (CPGI), the full clinical assessment protocol from which the PGSI short field screen derived. In the raw data one observes 2.8%, 1.5% and 0.5% classified as Low Risk, Moderate Risk and Problem Gambler, respectively, using the categories defined by Statistics Canada for the CCHS. Thus 4.8% are classified as “at risk.” After sample selection corrections these become 0.6%, 1.7% and 2.3%, respectively, or 4.6% in total. So virtually the same fraction are classified as “at risk,” but the composition is more heavily weighted toward those at greatest risk of a gambling disorder.
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 4 reflect 4,309,300 predictions.

These densities in Figure 4 allow one to see the average effects shown in Figure 3, the decrease in predicted Non-Indicated respondents from 0.946 to 0.839, 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 panel of Figure 4. We find that the predicted decrease in No Risk is statistically significant, in the sense that the 99th percentile of these p-values is 0.0001 or lower.29 Similarly, the average predicted increases in the Weakly Indicated, Moderately Indicated and Pathological Gambler categories (Figure 3) are also statistically significant, with the 99th percentile of p-values again being 0.001 lower in each case (Figure 4).

Figure 5 shows a decomposition of the processes underlying the sample selection correction, to better understand the logic. For each category of gambling problem or risk, it displays the conditional probability of being classified in that category depending on whether the subject is predicted to be “selected out” or “selected in” by the trigger question. For instance, if someone is predicted not to be selected in, the probability of them being classified as Weakly Indicated is 0.142; if that person is predicted to be selected in, the probability of them being classified as Weakly Indicated is 0.040. Since

29 The percentile value is purely descriptive, as a summary statistic for 43,093 p-values. The p-value is the inferential statistic.
the predicted probability of being selected in is 0.580, this implies that the weighted probability of being in the Weakly Indicated bin is \([0.580 \times 0.040] + [(1-0.580) \times 0.142] = 0.083\), which is the value shown in Figure 4 for being Weakly Indicated with sample selection correction.

Table 3 shows the predicted OR with respect to other psychiatric disorders for each category of the gambling hierarchy model with and without sample selection corrections. For each category of gambling problem or risk the OR for each disorder is much smaller when corrections are made for sample selection. Again, the upper bound of the 95% confidence interval with sample selection corrections is always below the lower bound of the same confidence interval without sample selection corrections.

2. Sample Selection Bias and Gambling Survey Screens

Survey screens have been traditionally designed to provisionally identify individuals who are likely to meet clinical criteria for GD. This has various implications for the design and format of the survey questions, which have evolved over time. Here we evaluate some of the issues that flow from that design objective as those relate to the use of trigger questions, ending with constructive suggestions to mitigate the sample selection biases such questions generate.

A. The Evolution of Trigger Questions

The history of the South Oaks Gambling Screen (SOGS) provides an important exemplar of these origins and concerns. The initial stages of the development of the instrument involved South Oaks Hospital patients already admitted for some alcohol or drug dependancy, and was prompted by knowledge from previous clinical treatment of the correlations between these addictions and gambling problems (Lesieur, Blume and Zoppa [1986]). In the initial pilots of screen designs, if “the patient denied any gambling, he or she was not interviewed further” (Lesieur and Blume [1987; p. 1185]). On
the other hand, later care and conversations might reveal that some deception had occurred, in which case the patient was re-interviewed (ibid.). The pilot questions, and the subsequent finalized SOGS, were directly motivated by the criteria stipulated in *DSM-III* (American Psychiatric Association [1987]), albeit with modifications to focus less on late stage, “desperation phase,” symptoms.\(^{30}\)

The final instrument, presented in Lesieur and Blume [1987; Appendix 1], was cross-validated by being given to 213 members of Gamblers Anonymous, 384 university students, and 152 hospital employees. The logic of this cross-validation was that the first group are self-identified as having gambling problems, while the last two groups were presumptively expected not to be DGs. Hence the detection of GD propensities of 98%, 5% and 1.3%, respectively, by the SOGS response scores was viewed as providing evidence of 2% false negatives, 5% *tentative* false positives, and 1.3% *tentative* false positives, respectively.

The clinical origins of SOGS did not mean that it automatically translated into an ideal epidemiological instrument, and indeed it was subsequently largely supplanted from that use by other instruments, such as the PGSI, thought to be more accurate. An important early warning was raised by one of the SOGS authors, Lesieur [1994], who carefully noted how seemingly minor changes in sampling procedures and question wording might completely change the interpretation, and claims of validity, of the instrument.

An important exception to the emphasis on clinical objectives for GD survey instruments is offered by Currie, Miller, Hodgins and Wang [2009], who argue that many gamblers who report no occurrent or historical gambling problem might be “at risk” in a broader public health sense. That is, someone identified in a survey as having no gambling problems might have a heightened propensity to engage in other behaviors that predict vulnerability to GD, and for that reason might be of interest to

\(^{30}\) The original *DSM-III* criteria stressed disruption of personal, family and employment activities. The revised criteria in *DSM-III-R* added physiological symptoms such as withdrawal problems.
There is no mention of a trigger question in the first epidemiological applications of SOGS in the United States reported by Volberg and Steadman [1988][1989], or in the revised SOGS surveys for New Zealand reported by Abbot and Volberg [1996]. One of the first surveys to have used a trigger question appears to be Dickerson, Baron, Hong and Cottrell [1996]. Since then, the use of trigger questions has become standard, particularly in large-scale epidemiological surveys, as the review of national prevalence studies by Williams, Volberg and Stevens [2012] shows. There are continuing debates about the nature of those trigger questions, but they generally concern whether participants should be asked about whether their gambling over lifetime or only past-year frames. There is also critical discussion about whether monetary loss thresholds should figure in questions. Stone et al. [2015] emphasize these issues, while also signaling awareness of potential sample selection bias introduced by use of trigger questions, but do not address measures to explicitly correct for it.

A somewhat aggressive defense of trigger questions is provided by the Australian Productivity Commission [1999; volume 3, page F14]:

The [Australian] National Gambling Survey did not administer the SOGS to all respondents – indeed there are good reasons why gambling surveys do not ask the problem gambling screen of all participants:

- questions about what people do when they gamble are clearly of no relevance to non gamblers. In the National Gambling Survey, respondents were classified as a non gambler only after they had answered 'no' to thirteen separate questions about whether they had participated in any of twelve specified gambling activities and an ‘any other’ gambling category. Hence, this detail of questioning should reliably identify a genuine non gambler.
- a problem gambling screen is of little or no relevance to infrequent gamblers because their gambling is very unlikely to be associated with problematic behaviour; but
- it is most appropriate to administer a problem gambling screen to those respondents whose gambling has a greater likelihood of giving rise to problems.

Indeed, as the NORC [National Opinion Research Center] study (Gerstein et al. 1999) noted: We chose to use these “filter” questions in the national survey after our pretesting indicated that nongamblers and very infrequent gamblers grew impatient with repeated questions about gambling-related problems (p. 19). For these reasons, the problem gambling instrument was administered only to that subset of gamblers considered most likely to experience problems related to their gambling – all ‘regular’ gamblers as defined by filter 2 and ‘big spending’ and other non-regular gamblers captured by filter 3.
We would rephrase the last sentence as follows: for these reasons, the GD screen used by Gerstein et al. [1999] was administered only to that subset of gamblers considered most likely on the basis of *ex ante* theory to experience gambling problems. As we discuss below, best-practice survey design *should* bring prior theory to bear, but for the purpose of gathering data that contributes to modeling sample exclusions, rather than as a basis for filtering out some information altogether.

The form of the trigger question is raised as an issue by Volberg and Williams [2012; p. 9] as follows:

A final important methodological variation that is known to have a significant impact on problem gambling prevalence rates concerns the threshold for administering problem gambling questions. Engaging in any gambling in the past year is a common criterion used to administer questions about problem gambling. However, Williams and Volberg [2009][2010] found that this criterion results in too many false positives on problem gambling screening instruments (as assessed by subsequent clinical assessment). These false positives can be significantly reduced by (a) using a higher threshold for the designation of problem gambling (i.e., CPGI$^{31} 5+$ versus CPGI 3+); and/or (b) requiring a minimal frequency of gambling in the past year (i.e., at least 10 times on some format) before administering problem gambling screens; and/or (c) resolving these cases of inconsistent gambling behaviour by automatically asking people to explain the discrepancy between their problem gambling classification in the absence of significant gambling behaviour, or intensive gambling involvement in the absence of reports of problems.

Indeed, Williams and Volberg [2009] conducted a careful evaluation of three survey administration features, and report disturbing effects on inferred GD prevalence:

- they found that just referring in the introduction to a “gambling survey” rather than a “health and recreation survey” caused a 133% increase in estimated GD prevalence;
- using face-to-face interviews rather than telephone interviews led to a 55% increase; and
- using a trigger question with a cutoff of C$300 in annual gambling losses, compared to the trigger of any gambling in the past year, would have implied a 42% decrease.

The conjectured rationale for the first effect is that gamblers like taking gambling surveys, which

---

$^{31}$ The citation, strictly speaking, refers to the PGSI, the short scored field screen of the CPGI.
economists regard as a classic sample selection effect. The second effect is simply demographic, and would be easy to correct with the right sample weights in the population: men respond more to one mode of interview than women, and men gamble much more than women. No explanation for the final effect is offered, although Williams and Volberg [2009; p.112] note that one of their subjects who was in this category revealed an interesting issue:

There was one individual with a CPGI score of 12 despite not reporting any past year gambling. It is interesting to note that this person reported having a history of problem gambling prior to the past 12 months, which may have influenced his responses to the CPGI past year questions.

This subject, it seems, had gambling under control in the year before the survey, but based on earlier history might be conjectured to still be vulnerable to GD under certain conditions. Such a fact might not be clinically important at point of presentation, but should be relevant to public health forecasting, or to regulatory officials deciding whether to license new gaming facilities.

Possible ambiguity of some threshold questions also raises sampling concerns. Blaszczynski, Dumlao and Lange [1997] cite evidence suggesting ambiguity in interpreting the question “How much money do you spend on gambling?” Over five case study vignettes considered by their subjects the most popular interpretation was the net amount of money spent in a session. But other subjects interpreted the same vignette in terms of initial stake, turnover, or even just losses, as well as some random responses disconnected to the information. Blaszczynski et al. [1997; p. 249ff.] suggest that the most relevant estimate of gambling expenditure is net expenditure. [...] It is recommended that future prevalence studies provide adequate instructions on how to calculate the net expenditure by drawing subjects’ attention to the difference between amounts invested and the residual at the conclusion of each session. It is suggested that wins reinvested during particular individual sessions should be ignored.

A similar issue was examined by Wood and Williams [2007], who evaluated 12 different ways of asking this question, and concluded (p. 72) that, “In general, retrospective estimates of gambling expenditures appear unreliable.” To be sure, some ways of asking the question elicited more reliable responses, by some sensible metrics. And it does not follow that other forms of detecting a gambling threshold
suffer the same ambiguities. For instance, asking if someone has gambled five times in the past year may be easier than asking them to tell you how much they spent on gambling in the past year, or even if they recall losing a certain amount of money in any one day in the past year.

There is widespread recognition of the difficulty of asking “how much money have you lost” questions. Some DGs erase prior losses within a gambling session from cognitive book-keeping as soon as they win; Rachlin [1990][2000] and Rachlin et al. [2015] argue that this is one of the characteristics that distinguishes DGs from self-controlled gamblers. Concern with this issue led Sharp et al. [2012], in a South African prevalence study, to pose the question as follows:

Thinking about the last time you participated in [ASK FOR EACH GAME EVER PLAYED, FROM A PREVIOUS QUESTION], approximately how much money would you say you staked on that occasion – that is the total amount in rands you put down to bet on that activity during that whole evening or day, not the amount you won and not the amount you ended up with at the end? Please take your time to think carefully about this.

They found that subjects identified as DGs based on their PGSI scores tended to take significantly longer to answer this question than people who reported regular gambling but did not score in the GD range.\(^{32}\)

### B. Mitigating the Effects of Trigger Questions

How might one mitigate some of the effects on prevalence estimates of survey screens 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, and indeed we have done that in Denmark as a result of the concerns

\(^{32}\) Sharp et al. [2012] further tried to encourage accurate responses by asking this question separately for each game the respondent reports playing. Since they know mean general house advantages for game types as set by South African regulations, this allows them to compute expected losses to the extent that subjects reported expenditures in the strict sense of that word. Of course this approach was profligate with subjects’ time, which can cause them to become impatient and consequently respond less accurately to questions in general.
identified here (see Harrison, Jessen, Lau and Ross [2016]). In this study, questions based on two different loss threshold quanta were asked of respondents at the end of the survey that was administered to all participants. This allowed analysis to compare the actual estimation of GD prevalence, across all levels in the hierarchy of risk, with the hypothetical estimates that would have been generated had those who failed to meet one threshold or the other been assigned to a “Non-Gambler” or “No Risk” category due to being excluded from further screening. We recognize that this can only be done if a few specific psychiatric disorders are the focus of the survey, given limitations on time needed for subject responses.

Several surveys have come close to this ideal, by employing extremely “light” trigger questions that only exclude from the sample those that have never engaged in any gambling over some period, including the mere purchase of a lottery ticket. One example is the British Gambling Prevalence Survey (BGPS) of 2010, which asked PGSI and DSM-IV questions for 73% of their entire sample of 7,756: see Wardle et al. [2011]. Figure 6 shows that although there are differences in prevalence when correcting for sample selection bias using the DSM-IV-based screen, they are not statistically significant or even quantitatively significant for policy purposes. The analysis of the BGPS also demonstrates that the sample selection correction does not always increase the fraction of the population predicted to be at risk: in this case that fraction drops from 5.0% to 4.8% with correction.

Another example of the value of asking an extremely light trigger question is the first wave of the Victorian Gambling Survey (VGS) of 2008, which asked PGSI questions for 75% of their entire sample of 15,000: see Billi et al. [2014][2015] and Stone et al. [2015]. Again, Figure 7 shows that although there are economically significant differences in prevalence overall when correcting for

---

33 The U.K. National Centre for Social Research, the Gambling Commission, and the UK Data Archive bear no responsibility for our analysis or interpretation of the BGPS. Figure B1 in Appendix B documents the claim about statistical insufficiency of the differences.

34 Stone et al. [2015] also present results from the Swedish Longitudinal Gambling Survey. The data from that study are not available for replication or review (Ulla Romild, Public Health Agency of Sweden; personal communication, October 23, 2016).
sample selection bias using the PGSI, the differences for the most important categories of Moderately Indicated and Pathological Gambler\(^{35}\) are not statistically significant\(^{36}\) or quantitatively significant for policy purposes. Of course, if the policy objective is to identify demographic slices that might be at risk, one would need to go beneath these aggregate population prevalence estimates to know if there is a sample selection bias.

On the other hand, the same VGS illustrates the risks of using additional threshold questions in order to reduce respondent time during the interview.\(^{37}\) Starting with the 75% of the sample that had gambled at all in the last 12 months, these surveyors added 1,057 individuals who had gambled before then, to arrive at a lifetime sample of gamblers of 81%. But they then employed an additional pre-screening procedure when applying the National Opinion Research Center DSM (NODS) screen of Gerstein et al. [1999] to measure lifetime gambling prevalence. This procedure asks 5 questions about gambling behavior, and only follows up with the additional questions of the full NODS instrument if someone responds affirmatively to one of those 5 pre-screening questions. This procedure drops the VGS sample by 11,075, so that we end up with only 8.5% of the full sample being evaluated with the NODS instrument. Unfortunately, this procedure is not statistically innocent: as Figure 8 shows, it leads to a statistically and quantitatively significant sample selection bias when inferring lifetime prevalence.

\(^{35}\) For consistency we repeat the categories of gambling problems and risk used in the NESARC data analysis (Figures 3, 4 and 5) rather than the categories reported by the original BGPS and VGS reports from the DSM-IV, PGSI, and NODS screens. In Figure 6 the original category is “Problem Gambling,” with a DSM-IV score of 3 or more. In Figure 7 the original categories are “Low Risk,” “Moderate Risk,” and “Problem Gambler,” respectively; as noted earlier, the PGSI uses “Problem Gambler” as synonymous with the DSM-IV’s “Pathological Gambler” (and, therefore, the DSM 5’s “Disordered Gambler.” In Figure 8 the original categories are “At Risk,” “Problem Gambler,” and “Pathological Gambler,” respectively.

\(^{36}\) Figure B2 in Appendix B documents the claim about statistical insignificance of the differences.

\(^{37}\) Another example of additional threshold questions being used is the Canadian Community Health Survey of Mental Health and Well-Being of 2002. Of the sample of 36,984, 24.6% said that they had not engaged in any of 13 gambling activities in the past year. Then 46.3% of the total sample was not asked the full set of CPGI questions because they had only gambled between 1 and 5 times, at most, for each of the 13 activities. And then 24.0% of the sample was not asked the full set of CPGI questions because they said that they were a non-gambler on the first CPGI question. There were 98 subjects that refused to answer the initial questions about gambling activity, resulting in only 1,759 being asked the full set of questions and having any chance of being scored as “at risk.” These deviations from the CPGI screen, and the PGSI index derived from it, were “approved by the authors of the scale” (Statistics Canada [2004; p.19]).
prevalence.\textsuperscript{38} The fraction of the Victorian adult population that is indicated as being at risk of GD jumps from 5.6\% to 10.5\% after correcting for sample selection bias, and the fraction of DGs increases from 0.9\% to 2.3\%.

Second, one can design surveys that use criteria for GD that do not rely on historical gambling experience to measure whether someone is at risk, as is the focus of virtually every trigger question we find in the literature. To the extent that one has a theoretically motivated structural model of the nexus of causal factors for GD, some of which may be present in the absence of gambling opportunities in a person’s environment, one can gain information about a respondent’s risk for developing GD that is independent of any historical gambling behavior meeting a threshold of frequency or financial loss. Note that in such modeling, the sense of “risk” is prospective, in contrast to the current risk of misdiagnosis that is operationalized in the PGSI and in DSM-based screens. The Focal Adult Gambling Screen (FLAGS) designed by Schellinck et al. \textsuperscript{[2015a][2015b]} is an instance of such a screen, which has been used by industry analysts around the world, but has been deployed by no prevalence study prior to the use of it by Harrison, Jessen, Lau and Ross \textsuperscript{[2016]} in Denmark. The value of prospective risk forecasting for policy around gambling facility licensing should be obvious, and this policy goal naturally complements the point we are making about using trigger questions that are not reliant on past gambling experience. When considering jurisdictions that have had bans on certain forms of gambling, or where the transactions costs of engaging in gambling have changed, it is quite possible that someone exhibits traits that would lead them to be at risk of gambling problems in different circumstances than they have experienced.\textsuperscript{39} Mitigation of sample selection bias and enhanced policy guidance might thus be achieved by the same research design strategy.

Third, where there is a need for some sort of trigger question or questions to avoid taking too

\textsuperscript{38} Figure B3 in Appendix B documents the claim about statistical significance of the differences.

\textsuperscript{39} To take a stark example, assume that gambling is illegal until one reaches a certain age of consent. Surveys of individuals who have just reached that age would show nobody at risk, but of course that says nothing about the future propensity of the individual to have gambling problems.
much time in surveys, one can build in random treatments to make it easier to identify sample selection bias. These treatments would be conditions that affect the likelihood of someone participating in a full survey, or engaging in deception due to sensitivity around a question. An example of the former would be financial incentives for participating in surveys, of the kind employed in some surveys and experiments. An example of the latter would be any one of a myriad of survey techniques for using “randomized response” methods to ensure that subjects are not revealing with certainty some sensitive information in response to a question.

Fourth, 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, whether or not they agree to participate. This is not a general option, since few non-Scandinavian countries have general registries, although recruiting from a Census may suffice if access to characteristics of the individuals recruited is possible.

3. Conclusions

Measurement of the population prevalence of the risk of gambling problems, and the psychiatric disorders with which they are correlated, play critical roles in public health policy and

---

40 For an example from surveys, consider the follow-up to the longitudinal Movement to Opportunity (MTO) field experiment, in which 30% of the sample was randomly assigned to more intensive follow-up; see Orr et al. [2003; Exhibit B, §B1.3] and DiNardo, McCrary and Sanbonmatsu [2006]. This randomized follow-up was in addition to the primary randomization to treatment: (i) a housing voucher with some strings attached and some counseling, (ii) a housing voucher with no strings attached and no counseling, and (iii) a control group. This additional randomization to more intensive follow-up had virtually no effect on results, since the effective response rates for the long-term MTO follow-up were around 90%, and similar across primary treatments. This methodological step was striking, since it provided some controlled basis for inferring sample attrition, which is formally identical to sample selection, albeit in the opposite direction (selecting out of the longitudinal sample). For an example from field experiments, see Harrison, Il and Lau [2014], where subjects were offered different non-risky incentives to participate and effects on measured risk aversion assessed after correcting for sample selection.

41 See Warner [1965] for the original idea, and Blair, Imai and Zhou [2015] for a recent review and application of variants. There are other non-randomized survey methods for encouraging truthful responses to sensitive questions.
policy around the licensing of casinos and other gambling facilities. It should make a substantive
difference to policy assessment and forecasts of consequence whether the fraction “at risk” of being
current DGs is 1.3% or 7.7%, and that is the pure effect of allowing for sample selection bias in the
application of a major, widely-cited, and conventional survey of the U.S. population. In fact, one
conjecture as to why investigation of GD prevalence was dropped from follow-up waves of the
NESARC in the United States and from the Mental Health module of the Canadian Community
Health Survey is that the uncorrected population prevalence was too low to justify resources and
interview time asking the questions. Hence it becomes “settled” belief that the prevalence of GD is
tiny, and no data are ever then collected to question that belief. It would represent a shameful failure
of linkage between best research practice and best policy design if these statistical biases in
measurement of population prevalence drove substantive decisions concerning the regulation of
gambling or the allocation of resources toward the treatment of GD.

One immediate substantive implication of our findings is to ask if comparable biases distort
inferences about other psychiatric disorders, since prevalence surveys for every major psychiatric
disorder typically use comparable trigger questions. Although it is conceivable that the distortion could
be in any direction, our a priori expectation would be that the distortions lead to understatements of
prevalence across the board, given the sensitive nature of the trigger questions.
Table 1: Logistic Odds Ratio Calculations for Pathological Gambling

<table>
<thead>
<tr>
<th></th>
<th>Odds Ratio</th>
<th>Standard Error</th>
<th>$t$ for $H_0: OR = 1$</th>
<th>$p$-value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Total Effects with Logistic Model and No Covariates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>6.1</td>
<td>0.52</td>
<td>21.1</td>
<td>&lt; 0.001</td>
<td>5.1</td>
</tr>
<tr>
<td>Drug</td>
<td>5.2</td>
<td>0.61</td>
<td>14.3</td>
<td>&lt; 0.001</td>
<td>4.2</td>
</tr>
<tr>
<td>Nicotine</td>
<td>7.0</td>
<td>0.62</td>
<td>22.0</td>
<td>&lt; 0.001</td>
<td>5.8</td>
</tr>
<tr>
<td>Mood</td>
<td>4.8</td>
<td>0.44</td>
<td>17.3</td>
<td>&lt; 0.001</td>
<td>4.0</td>
</tr>
<tr>
<td>Anxiety</td>
<td>3.4</td>
<td>0.32</td>
<td>13.1</td>
<td>&lt; 0.001</td>
<td>2.8</td>
</tr>
<tr>
<td>Personality</td>
<td>9.0</td>
<td>0.81</td>
<td>24.2</td>
<td>&lt; 0.001</td>
<td>7.4</td>
</tr>
<tr>
<td><strong>B. Marginal Effects with Logistic Model and No Covariates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>2.6</td>
<td>0.19</td>
<td>13.1</td>
<td>&lt; 0.001</td>
<td>2.2</td>
</tr>
<tr>
<td>Drug</td>
<td>1.3</td>
<td>0.14</td>
<td>2.2</td>
<td>0.03</td>
<td>1.0</td>
</tr>
<tr>
<td>Nicotine</td>
<td>2.7</td>
<td>0.21</td>
<td>13.2</td>
<td>&lt; 0.001</td>
<td>2.4</td>
</tr>
<tr>
<td>Mood</td>
<td>1.7</td>
<td>0.21</td>
<td>4.2</td>
<td>&lt; 0.001</td>
<td>1.3</td>
</tr>
<tr>
<td>Anxiety</td>
<td>1.1</td>
<td>0.13</td>
<td>0.9</td>
<td>0.39</td>
<td>0.9</td>
</tr>
<tr>
<td>Personality</td>
<td>4.1</td>
<td>0.65</td>
<td>8.8</td>
<td>&lt; 0.001</td>
<td>2.9</td>
</tr>
<tr>
<td><strong>C. Total Effects with Logistic Model and Covariates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>5.2</td>
<td>0.48</td>
<td>17.8</td>
<td>&lt; 0.001</td>
<td>4.3</td>
</tr>
<tr>
<td>Drug</td>
<td>3.9</td>
<td>0.45</td>
<td>11.9</td>
<td>&lt; 0.001</td>
<td>3.1</td>
</tr>
<tr>
<td>Nicotine</td>
<td>6.1</td>
<td>0.51</td>
<td>21.4</td>
<td>&lt; 0.001</td>
<td>5.1</td>
</tr>
<tr>
<td>Mood</td>
<td>5.1</td>
<td>0.52</td>
<td>16.3</td>
<td>&lt; 0.001</td>
<td>4.2</td>
</tr>
<tr>
<td>Anxiety</td>
<td>3.8</td>
<td>0.35</td>
<td>14.5</td>
<td>&lt; 0.001</td>
<td>3.2</td>
</tr>
<tr>
<td>Personality</td>
<td>7.8</td>
<td>0.68</td>
<td>23.6</td>
<td>&lt; 0.001</td>
<td>6.6</td>
</tr>
</tbody>
</table>
### D. Marginal Effects with Logistic Model and Covariates

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol</td>
<td>2.2</td>
<td>0.18</td>
<td>9.4</td>
<td>&lt; 0.001</td>
<td>1.8</td>
<td>2.6</td>
</tr>
<tr>
<td>Drug</td>
<td>1.1</td>
<td>0.13</td>
<td>1.2</td>
<td>0.24</td>
<td>0.9</td>
<td>1.4</td>
</tr>
<tr>
<td>Nicotine</td>
<td>2.8</td>
<td>0.22</td>
<td>13.5</td>
<td>&lt; 0.001</td>
<td>2.4</td>
<td>3.3</td>
</tr>
<tr>
<td>Mood</td>
<td>1.9</td>
<td>0.26</td>
<td>5.0</td>
<td>&lt; 0.001</td>
<td>1.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Anxiety</td>
<td>1.3</td>
<td>0.15</td>
<td>2.1</td>
<td>0.04</td>
<td>1.0</td>
<td>1.6</td>
</tr>
<tr>
<td>Personality</td>
<td>3.7</td>
<td>0.58</td>
<td>8.4</td>
<td>&lt; 0.001</td>
<td>2.7</td>
<td>5.1</td>
</tr>
</tbody>
</table>

### E. Marginal Odds Ratios for Demographic Covariates with Logistic Model

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.4</td>
<td>0.04</td>
<td>-9.3</td>
<td>&lt; 0.001</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Black</td>
<td>3.1</td>
<td>0.25</td>
<td>14.1</td>
<td>&lt; 0.001</td>
<td>2.7</td>
<td>3.7</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.9</td>
<td>0.11</td>
<td>-1.2</td>
<td>0.24</td>
<td>0.7</td>
<td>1.1</td>
</tr>
<tr>
<td>Age 30-44</td>
<td>0.9</td>
<td>0.17</td>
<td>-0.3</td>
<td>0.77</td>
<td>0.7</td>
<td>1.4</td>
</tr>
<tr>
<td>Age 45-64</td>
<td>1.3</td>
<td>0.22</td>
<td>1.7</td>
<td>0.09</td>
<td>0.9</td>
<td>1.8</td>
</tr>
<tr>
<td>Age 65+</td>
<td>1.1</td>
<td>0.18</td>
<td>0.7</td>
<td>0.51</td>
<td>0.8</td>
<td>1.5</td>
</tr>
<tr>
<td>Separated</td>
<td>1.5</td>
<td>0.15</td>
<td>4.3</td>
<td>&lt; 0.001</td>
<td>1.2</td>
<td>1.8</td>
</tr>
<tr>
<td>Midwest</td>
<td>1.1</td>
<td>0.14</td>
<td>0.8</td>
<td>0.45</td>
<td>0.8</td>
<td>1.4</td>
</tr>
<tr>
<td>South</td>
<td>0.9</td>
<td>0.10</td>
<td>-0.9</td>
<td>0.35</td>
<td>0.7</td>
<td>1.1</td>
</tr>
<tr>
<td>West</td>
<td>1.7</td>
<td>0.22</td>
<td>4.0</td>
<td>&lt; 0.001</td>
<td>1.3</td>
<td>2.2</td>
</tr>
<tr>
<td>High School</td>
<td>0.9</td>
<td>0.13</td>
<td>-0.1</td>
<td>0.91</td>
<td>0.7</td>
<td>1.3</td>
</tr>
<tr>
<td>Some College</td>
<td>1.0</td>
<td>0.08</td>
<td>-0.3</td>
<td>0.78</td>
<td>0.8</td>
<td>1.1</td>
</tr>
<tr>
<td>College</td>
<td>0.5</td>
<td>0.07</td>
<td>-5.3</td>
<td>&lt; 0.001</td>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>Graduate</td>
<td>0.7</td>
<td>0.13</td>
<td>-2.2</td>
<td>0.03</td>
<td>0.4</td>
<td>0.9</td>
</tr>
<tr>
<td>Income 2</td>
<td>0.9</td>
<td>0.12</td>
<td>-0.4</td>
<td>0.72</td>
<td>0.7</td>
<td>1.2</td>
</tr>
<tr>
<td>Income 3</td>
<td>0.6</td>
<td>0.06</td>
<td>-5.3</td>
<td>&lt; 0.001</td>
<td>0.5</td>
<td>0.7</td>
</tr>
<tr>
<td>Income 4</td>
<td>1.5</td>
<td>0.17</td>
<td>3.3</td>
<td>0.002</td>
<td>1.2</td>
<td>1.9</td>
</tr>
</tbody>
</table>
Table 2: Semi-Nonparametric Odds Ratio Calculations for Pathological Gambling

<table>
<thead>
<tr>
<th></th>
<th>Odds Ratio</th>
<th>Standard Error</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Total Effects with No Covariates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>6.0</td>
<td>&lt; 0.001</td>
<td>6.0</td>
</tr>
<tr>
<td>Drug</td>
<td>6.2</td>
<td>&lt; 0.001</td>
<td>6.2</td>
</tr>
<tr>
<td>Nicotine</td>
<td>6.7</td>
<td>&lt; 0.001</td>
<td>6.7</td>
</tr>
<tr>
<td>Mood</td>
<td>4.8</td>
<td>&lt; 0.001</td>
<td>4.8</td>
</tr>
<tr>
<td>Anxiety</td>
<td>4.2</td>
<td>&lt; 0.001</td>
<td>4.2</td>
</tr>
<tr>
<td>Personality</td>
<td>8.2</td>
<td>&lt; 0.001</td>
<td>8.2</td>
</tr>
<tr>
<td>B. Marginal Effects with No Covariates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>4.2</td>
<td>0.004</td>
<td>4.2</td>
</tr>
<tr>
<td>Drug</td>
<td>1.6</td>
<td>0.001</td>
<td>1.6</td>
</tr>
<tr>
<td>Nicotine</td>
<td>3.3</td>
<td>0.003</td>
<td>3.3</td>
</tr>
<tr>
<td>Mood</td>
<td>2.4</td>
<td>0.002</td>
<td>2.4</td>
</tr>
<tr>
<td>Anxiety</td>
<td>2.3</td>
<td>0.001</td>
<td>2.3</td>
</tr>
<tr>
<td>Personality</td>
<td>4.6</td>
<td>0.005</td>
<td>4.6</td>
</tr>
</tbody>
</table>
Figure 1: Comorbidity of Gambling Disorder and Other Psychiatric Disorders

Estimated Odds Ratios using Semi-Nonparametric Ordered Response Model
Source: National Epidemiological Survey on Alcohol and Related Conditions (NESARC)

- Alcohol
  - Total: 4.25
  - Marginal: 6.05
- Drugs
  - Total: 1.6
  - Marginal: 6.17
- Nicotine
  - Total: 3.27
  - Marginal: 6.69
- Mood
  - Total: 2.4
  - Marginal: 4.82
- Anxiety
  - Total: 2.35
  - Marginal: 4.24
- Personality
  - Total: 4.65
  - Marginal: 8.16

Odds Ratio
Figure 2: Effect of Sample Selection Correction on Estimates of Comorbidity Marginal Effects with Gambling Disorder

Estimated Odds Ratios using Semi-Nonparametric Ordered Response Model
Source: National Epidemiological Survey on Alcohol and Related Conditions (NESARC)

<table>
<thead>
<tr>
<th>Comorbidity</th>
<th>Uncorrected</th>
<th>Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol</td>
<td>1.18</td>
<td>4.25</td>
</tr>
<tr>
<td>Drugs</td>
<td>1.6</td>
<td>1.12</td>
</tr>
<tr>
<td>Nicotine</td>
<td>1.26</td>
<td>3.27</td>
</tr>
<tr>
<td>Mood</td>
<td>1.47</td>
<td>2.4</td>
</tr>
<tr>
<td>Anxiety</td>
<td>1.34</td>
<td>2.35</td>
</tr>
<tr>
<td>Personality</td>
<td>1.75</td>
<td>4.65</td>
</tr>
</tbody>
</table>
Table 3: Semi-Nonparametric Odds Ratio Calculations for Gambling Risks

Each standard error is less than 0.002, so 95% confidence intervals are the same

<table>
<thead>
<tr>
<th></th>
<th>Odds Ratio with No Sample Selection Correction</th>
<th>Odds Ratio with No Sample Selection Correction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Weakly Indicated</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>1.95</td>
<td>1.02</td>
</tr>
<tr>
<td>Drug</td>
<td>1.25</td>
<td>1.02</td>
</tr>
<tr>
<td>Nicotine</td>
<td>1.72</td>
<td>1.03</td>
</tr>
<tr>
<td>Mood</td>
<td>1.50</td>
<td>1.05</td>
</tr>
<tr>
<td>Anxiety</td>
<td>1.48</td>
<td>1.04</td>
</tr>
<tr>
<td>Personality</td>
<td>2.01</td>
<td>1.07</td>
</tr>
<tr>
<td><strong>B. Moderately Indicated</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>2.91</td>
<td>1.10</td>
</tr>
<tr>
<td>Drug</td>
<td>1.42</td>
<td>1.06</td>
</tr>
<tr>
<td>Nicotine</td>
<td>2.40</td>
<td>1.14</td>
</tr>
<tr>
<td>Mood</td>
<td>1.91</td>
<td>1.24</td>
</tr>
<tr>
<td>Anxiety</td>
<td>1.88</td>
<td>1.18</td>
</tr>
<tr>
<td>Personality</td>
<td>3.08</td>
<td>1.35</td>
</tr>
<tr>
<td><strong>C. Pathological Gambler</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>4.25</td>
<td>1.18</td>
</tr>
<tr>
<td>Drug</td>
<td>1.60</td>
<td>1.12</td>
</tr>
<tr>
<td>Nicotine</td>
<td>3.27</td>
<td>1.26</td>
</tr>
<tr>
<td>Mood</td>
<td>2.40</td>
<td>1.47</td>
</tr>
<tr>
<td>Anxiety</td>
<td>2.35</td>
<td>1.34</td>
</tr>
<tr>
<td>Personality</td>
<td>4.65</td>
<td>1.75</td>
</tr>
</tbody>
</table>
Figure 3: Predicted Prevalence of Gambling Risk With and Without Sample Selection Correction

Estimated Probabilities using Semi-Nonparametric Ordered Response Model
Source: National Epidemiological Survey on Alcohol and Related Conditions (NESARC)

Non-Indicated
- Uncorrected: 0.946
- Corrected: 0.839

Weakly Indicated
- Uncorrected: 0.040
- Corrected: 0.083

Moderately Indicated
- Uncorrected: 0.009
- Corrected: 0.039

Pathological Gambler
- Uncorrected: 0.004
- Corrected: 0.038

Predicted Probability
Figure 4: Statistical Significance of Sample Selection Corrections for Gambling Risk

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

- Non-Indicated
- Weakly Indicated
- Moderately Indicated
- Pathological Gambler

90th percentile p-value < 0.001
95th percentile p-value < 0.001
99th percentile p-value < 0.001
Figure 5: Predicted Probability of Gambling Risk Conditional on Sample Selection or Not

Predicted Probability of Sample Selection = 0.58
Estimated Probabilities using Semi-Nonparametric Ordered Response Model
Source: National Epidemiological Survey on Alcohol and Related Conditions (NESARC)

Non-Indicated
- No Selection
  - Selection
    - 0.692
    - 0.947

Weakly Indicated
- No Selection
  - Selection
    - 0.142
    - 0.040

Moderately Indicated
- No Selection
  - Selection
    - 0.081
    - 0.009

Pathological Gambler
- No Selection
  - Selection
    - 0.085
    - 0.004
Figure 6: Predicted Prevalence of Gambling Disorders in the U.K., Measured with the DSM-IV Screen, With and Without Sample Selection Correction

Estimated Probabilities using Semi-Nonparametric Ordered Response Model
Source: British Gambling Prevalence Survey of 2010
Fraction at any risk level changes from 0.050 to 0.048 with correction

Weakly Indicated
Uncorrected 0.041
Corrected 0.035

Moderately Indicated
Uncorrected 0.005
Corrected 0.004

Pathological Gambler
Uncorrected 0.004
Corrected 0.008
Figure 7: Predicted Prevalence of Gambling Disorders in Victoria (Australia), Measured with the PGSI Screen, With and Without Sample Selection Correction

Estimated Probabilities using Semi-Nonparametric Ordered Response Model
Fraction at any risk level changes from 0.088 to 0.116 with correction

- Weakly Indicated
  - Uncorrected: 0.024
  - Corrected: 0.030

- Moderately Indicated
  - Uncorrected: 0.057
  - Corrected: 0.079

- Pathological Gambler
  - Uncorrected: 0.007
  - Corrected: 0.007

Predicted Probability
Figure 8: Predicted Prevalence of Gambling Disorders in Victoria (Australia), Measured with the NODS Screen, With and Without Sample Selection Correction

Estimated Probabilities using Semi-Nonparametric Ordered Response Model
Fraction at any risk level changes from 0.056 to 0.105 with correction


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


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.


Chucro, Lucian; Tace, Marianna; Johnson, Robert; Buie, Tracy; Hill, Mary Ann; Volberg, Rachel; Harwood, Henrik; Tucker, Adam; Christiansen, Eugene; Cummings, Will, and Sinclair, Sebastian, *Gambling Impact and Behavior Study: Report to the National Gambling Impact Study Commission* (Chicago: National Opinion Research Center at the University of Chicago, 1999).


Maddala, G. S., Limited-Dependent and Qualitative Variables in Econometrics (New York: Cambridge University Press, 1983).


Nower, Lia; Martins, Silvia; Lin, Keng-Han, and Blanco, Carlos, “Subtypes of Disordered Gamblers: Results from the National Epidemiologic Survey on Alcohol and Related Conditions,” Addiction 108(4), 2013, 789-798.


Petry, Nancy; Blanco, Carlos; Jin, Chelsea, and Grant, Bridget F., “Concordance Between Gambling Disorder Diagnoses in the DSM-IV and DSM-5: Results from the National Epidemiologic Survey on Alcohol and Related Disorders [sic],” Psychology of Addictive Behaviors, 28(2), 2014, 586-591.

Pietrzak, Robert H.; Morasco, Banjamin J.; Blanco, Carlos; Grant, Bridget F., and Petry, Nancy M., “Gambling Level and Psychiatric and Medical Disorders in Older Adults: Results from the National Epidemiologic Survey on Alcohol and Related Conditions,” American Journal of Geriatric Psychiatry 15, 2007, 301-313.


Sharp, Carla; Steinberg, Lynne; Yaroslavsky, Ilya; Hofmeyr, Andre; Dellis, Andrew; Ross, Don, and Kincaid, Harold, “An Item Response Theory Analysis of the Problem Gambling Severity Index,” *Assessment*, 19(2), 2012, 167-175.


Stone, Christine; Romild, Ulla; Abbott, Max; Young, Kristal; Billi, Rosa, and Volberg, Rachel, “Effects of Different Screening and Scoring Thresholds on PGSI Gambling Risk Segments,” *International Journal of Mental Health and Addiction* 13, 2015, 82-102.


Wardle, Heather; Moody, Alison; Spence, Suzanne; Orford, Jim; Volberg, Rachel; Jotangia, Dhriti; Griffiths, Mark; Hussey, David, and Dobbie, Fiona, *British Gambling Prevalence Survey 2010* (London, UK: National Centre for Social Research, 2011).


Appendix A: Classifying Pathological Gamblers in NESARC (NOT FOR PUBLICATION)

The NESARC provides information on responses to 15 questions about gambling, which map into the 10 diagnostic criteria of DSM-IV. There appears to be no public documentation about how that mapping occurs, although it does seem “obvious” on an a priori basis, as explained below. On the other hand, there is an apparent mis-match between the mapped classifications from the raw data on these 15 questions and the “summary classification” of pathological gambling provided by the NESARC data files. This is of some significance because it appears that every researcher that has used the NESARC data has used this summary classification, rather than re-generating it from the ground up.

That summary measure is provided in section 13 of the data file codebook:

We focus here on lifetime questions, although the same issues apply to questions about the last 12 months.

The 15 NESARC questions are displayed on the next two pages. Pay attention to the Check Item 12.1 at the bottom of section 2, containing these 15 questions:

What this means is that someone is declared to be PG if they answer 5 of the 15 NESARC questions, not if they meet 10 of the DSM-IV criteria that these 15 question responses map into. This is in direct conflict with how the PG classification is defined in the papers reporting data from the NESARC. For example, Petry, Stinson and Grant [2005; p. 567] say that “Lifetime AUDADIS-IV diagnoses of pathological gambling required the respondent to meet at least 5 of the 10 DSM-IV criteria in the 12-month period preceding the interview and/or before that 12-month period. Fifteen symptom items operationalized the 10 pathological gambling criteria.”
Section 12 – BETTING

(Show Flashcard 7)

Statement:
Now I’d like to ask you a few questions about gambling. By gambling I mean playing cards for money, betting on the horses or dogs or sports games, playing the stock or commodities market, buying lottery tickets or playing bingo or KENO or gambling at a casino, including playing the slot machines.

1. Have you ever gambled at least 5 times in any one year of your life?
   1. Yes
   2. No - Skip to Section 13, page 126

2a. The next few questions are about experiences that people have had with gambling. As I read each experience, please tell me if it has EVER happened to you.

   In your ENTIRE LIFE did you EVER... (Pause)

(Repeat phrase frequently)

<table>
<thead>
<tr>
<th>Experience</th>
<th>Yes</th>
<th>No</th>
<th>Mark Yes in column c</th>
<th>Yes</th>
<th>No</th>
<th>Mark Yes in column c</th>
<th>Yes</th>
<th>No</th>
<th>Mark Yes in column c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamble to get out of a bad mood – like feeling nervous, sad or down?</td>
<td>1.</td>
<td>2. No - Go to next experience</td>
<td></td>
<td></td>
<td>1.</td>
<td>2. No - Mark Yes in column c</td>
<td></td>
<td></td>
<td>1.</td>
</tr>
<tr>
<td>Gamble to forget your problems?</td>
<td>1.</td>
<td>2. No - Go to next experience</td>
<td>1.</td>
<td>2. No - Mark Yes in column c</td>
<td>1.</td>
<td>2. No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More than once try to quit or cut down on your gambling, but found you couldn’t do it?</td>
<td>1.</td>
<td>2. No - Go to next experience</td>
<td>1.</td>
<td>2. No - Mark Yes in column c</td>
<td>1.</td>
<td>2. No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Find that you had to increase the amount of money you would gamble to keep it exciting?</td>
<td>1.</td>
<td>2. No - Go to next experience</td>
<td>1.</td>
<td>2. No - Mark Yes in column c</td>
<td>1.</td>
<td>2. No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spend a lot of time gambling, planning your bets or studying the odds?</td>
<td>1.</td>
<td>2. No - Go to next experience</td>
<td>1.</td>
<td>2. No - Mark Yes in column c</td>
<td>1.</td>
<td>2. No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spend a lot of time thinking about ways to get money together so you could gamble?</td>
<td>1.</td>
<td>2. No - Go to next experience</td>
<td>1.</td>
<td>2. No - Mark Yes in column c</td>
<td>1.</td>
<td>2. No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spend a lot of time thinking about the times when you won or lost?</td>
<td>1.</td>
<td>2. No - Go to next experience</td>
<td>1.</td>
<td>2. No - Mark Yes in column c</td>
<td>1.</td>
<td>2. No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have job or school trouble because of your gambling – like missing too much work, being demoted at work, losing your job or dropping out of school?</td>
<td>1.</td>
<td>2. No - Go to next experience</td>
<td>1.</td>
<td>2. No - Mark Yes in column c</td>
<td>1.</td>
<td>2. No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Break up or come close to breaking up with anyone who was important to you because of your gambling?</td>
<td>1.</td>
<td>2. No - Go to next experience</td>
<td>1.</td>
<td>2. No - Mark Yes in column c</td>
<td>1.</td>
<td>2. No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Try to keep your family or friends from knowing how much you gambled?</td>
<td>1.</td>
<td>2. No - Go to next experience</td>
<td>1.</td>
<td>2. No - Mark Yes in column c</td>
<td>1.</td>
<td>2. No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have such financial trouble as a result of your gambling that you had to get help with living expenses from family, friends or welfare?</td>
<td>1.</td>
<td>2. No - Go to next experience</td>
<td>1.</td>
<td>2. No - Mark Yes in column c</td>
<td>1.</td>
<td>2. No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Find that you became restless, irritable or anxious when trying to quit or cut down on your gambling?</td>
<td>1.</td>
<td>2. No - Go to next experience, page 124</td>
<td>1.</td>
<td>2. No - Mark Yes in column c</td>
<td>1.</td>
<td>2. No</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- A2 -
As it happens, it is relatively easy to see how the 15 questions map into the 10 DSM-IV criteria. From page 618 of the DSM-IV Manual, these 10 criteria are listed below, and Table A1 shows the 15 NESARC questions and the DSM criteria number they map into.

### Diagnostic criteria for 312.31 Pathological Gambling

A. Persistent and recurrent maladaptive gambling behavior as indicated by five (or more) of the following:

1. is preoccupied with gambling (e.g., preoccupied with reliving past gambling experiences, handicapping or planning the next venture, or thinking of ways to get money with which to gamble)
2. needs to gamble with increasing amounts of money in order to achieve the desired excitement
3. has repeated unsuccessful efforts to control, cut back, or stop gambling
4. is restless or irritable when attempting to cut down or stop gambling
5. gambles as a way of escaping from problems or of relieving a dysphoric mood (e.g., feelings of helplessness, guilt, anxiety, depression)
6. after losing money gambling, often returns another day to get even ("chasing" one's losses)
7. lies to family members, therapist, or others to conceal the extent of involvement with gambling
8. has committed illegal acts such as forgery, fraud, theft, or embezzlement to finance gambling
9. has jeopardized or lost a significant relationship, job, or educational or career opportunity because of gambling
10. relies on others to provide money to relieve a desperate financial situation caused by gambling

B. The gambling behavior is not better accounted for by a Manic Episode.
<table>
<thead>
<tr>
<th>NESARC Gambling Questions</th>
<th>DSM Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Gamble to get out of a bad mood – like feeling nervous, sad or down?</td>
<td>5</td>
</tr>
<tr>
<td>(2) Gamble to forget your problems?</td>
<td>5</td>
</tr>
<tr>
<td>(3) More than once try to quit or cut down on your gambling, but found you couldn’t do it?</td>
<td>3</td>
</tr>
<tr>
<td>(4) Find that you had to increase the amount of money you would gamble to keep it exciting?</td>
<td>2</td>
</tr>
<tr>
<td>(5) Spend a lot of time gambling, planning your bets or studying the odds?</td>
<td>1</td>
</tr>
<tr>
<td>(6) Spend a lot of time thinking about ways to get money together so you could gamble?</td>
<td>1</td>
</tr>
<tr>
<td>(7) Spend a lot of time thinking about the times when you won or lost?</td>
<td>1</td>
</tr>
<tr>
<td>(8) Have job or school trouble because of your gambling – like missing too much work, being demoted at work, losing your job or dropping out of school?</td>
<td>9</td>
</tr>
<tr>
<td>(8a) Break up or come close to breaking up with anyone who was important to you because of your gambling?</td>
<td>9</td>
</tr>
<tr>
<td>(9) Try to keep your family or friends from knowing how much you gamble?</td>
<td>7</td>
</tr>
<tr>
<td>(10) Have such financial trouble as a result of your gambling that you had to get help with living expenses from family, friends or welfare?</td>
<td>10</td>
</tr>
<tr>
<td>(11) Find that you become restless, irritable or anxious when trying to quit or cut down in your gambling?</td>
<td>4</td>
</tr>
<tr>
<td>(12) Raise gambling money by writing a bad check, signing someone else’s name to a check, stealing, cashing someone else’s check or in some other illegal way?</td>
<td>8</td>
</tr>
<tr>
<td>(13) Find that you had to gamble again as soon as possible after LOSING in order to win back your losses?</td>
<td>6</td>
</tr>
<tr>
<td>(14) Find that you had to gamble again as soon as possible after WINNING in order to win more?</td>
<td>2</td>
</tr>
</tbody>
</table>
If one undertakes this mapping for each individual in the survey, there are actually 207 PG subjects in these data, not 195 as tabulated in section 14 of the data. Here is a tabulation:

<table>
<thead>
<tr>
<th>Number of PG DSM-IV criteria met</th>
<th>pg_lifetime</th>
<th>pg_lifetime met</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>40,814</td>
<td>0</td>
<td>40,814</td>
</tr>
<tr>
<td>1</td>
<td>1,251</td>
<td>0</td>
<td>1,251</td>
</tr>
<tr>
<td>2</td>
<td>448</td>
<td>0</td>
<td>448</td>
</tr>
<tr>
<td>3</td>
<td>257</td>
<td>0</td>
<td>257</td>
</tr>
<tr>
<td>4</td>
<td>107</td>
<td>9</td>
<td>116</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>59</td>
<td>78</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>42</td>
<td>43</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>39</td>
<td>40</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>42,898</td>
<td>195</td>
<td>43,093</td>
</tr>
</tbody>
</table>

We see that 9 individuals were incorrectly defined to be PG when they only met 4 of the DSM-IV criteria, but that 21 individuals were defined as not being PG when they did indeed meet 5 or more of the DSM-IV. We do not know that all of these 21 were incorrectly defined, since we need to decide whether or not to check the DSM-IV exclusionary criterion, that they did not suffer from Manic Episodes.

As it happens, the NESARC has information on that criteria, and 3 of these 21 did exhibit such Manic Episodes if we choose to implement this exclusionary criteria using these NESARC data. We do not rule out cases associated with recent illness or bereavement. If we drop anyone with such episodes, we see something more problematic:

```
.tab pg_dsm_count pg_lifetime if manic_episodes==0, missing
```

<table>
<thead>
<tr>
<th>Number of PG DSM-IV criteria met</th>
<th>pg_lifetime</th>
<th>pg_lifetime met</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>39,515</td>
<td>0</td>
<td>39,515</td>
</tr>
<tr>
<td>1</td>
<td>1,159</td>
<td>0</td>
<td>1,159</td>
</tr>
<tr>
<td>2</td>
<td>401</td>
<td>0</td>
<td>401</td>
</tr>
<tr>
<td>3</td>
<td>218</td>
<td>0</td>
<td>218</td>
</tr>
<tr>
<td>4</td>
<td>92</td>
<td>5</td>
<td>97</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>45</td>
<td>61</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>41,403</td>
<td>142</td>
<td>41,545</td>
</tr>
</tbody>
</table>

-A5-
In other words, many of the 195 originally classified as PG should have been excluded. Specifically, 53 subjects:

```
. tab pg_lifetime manic_episodes, missing
```

<table>
<thead>
<tr>
<th></th>
<th>manic_episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>41,403</td>
</tr>
<tr>
<td>Yes</td>
<td>142</td>
</tr>
<tr>
<td>Total</td>
<td>41,545</td>
</tr>
</tbody>
</table>

We do not endorse this way of implementing the DSM-IV exclusionary criteria, but simply note here the consequences for sample size.

The hypothesis about what was done to generate the summary measure of PG that appears to have been used by all researchers using NESERC is confirmed if one does a “count” of the number of NESARC questions answered affirmatively:

```
. tab pg_count pg_lifetime, missing
```

<table>
<thead>
<tr>
<th>Number of PG NESARC criteria met</th>
<th>pg_lifetime</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>40,814</td>
</tr>
<tr>
<td>1</td>
<td>1,066</td>
</tr>
<tr>
<td>2</td>
<td>486</td>
</tr>
<tr>
<td>3</td>
<td>257</td>
</tr>
<tr>
<td>4</td>
<td>133</td>
</tr>
<tr>
<td>5</td>
<td>77</td>
</tr>
<tr>
<td>6</td>
<td>46</td>
</tr>
<tr>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>42,898</td>
</tr>
</tbody>
</table>

So we see that an individual was only classified as being PG in the summary measure if they had 5 of the 15 NESARC criteria. The fact that researchers used this summary measure is apparent from inspection of the reported sample size of 195 PG (e.g., Petry et al. [2005; Table 1, p. 568] or Blanco et al. [2006; Table 1, p.947]).

On the other hand, the above tabulation poses a puzzle, since there are many individuals with 5 or more of the 15 NESARC criteria that are not classified as PG. In fact, there are 77+46+13+5+1 = 142 such individuals. It is not as if these 142 individuals met some of the NESARC criteria that do not map into DSM criteria, since the NESARC criteria that they meet span all 15 NESARC questions. There appears to have been some algorithm used to defined PG in the summary measure that is not documented. For this reason we prefer to re-generate our own measure of PG using the DSM-IV criteria.

There are 207 correct DSM-IV based PG individuals in the NESARC: -A6-
There have been some variants of the DSM-IV threshold criteria applied in some research. For instance, as noted in the text, Fisher [2000] prefers the term “problem” to “pathological” outside of a shared, clinical diagnosis, and then (p. 33) defines problem gamblers using either 5 or more “yes” responses to the DSM-IV criteria or 3 or 4 “yes” responses that include at least one from items 8, 9 or 10 of the criteria. These last three refer to the adverse consequences of gambling. This definition can only logically increase the number of “classified” gamblers compared to the use of 5 or more “yes” responses. In fact, for the NESARC data this results in 256 “classified” gamblers compared to the 207 correct DSM-IV PG individuals:

It is also worthwhile doing the correct analysis using the DSM-V criteria. The new criteria drop the “illegal acts” question, and require that only 4 criteria be met. In this respect we mimic the approach of Petry, Blanco, Jon and Grant [2014] to “re-cycling” the NESARC data in this manner:
So we now have 321 individuals with PG, using the DSM-V criteria.

We contacted researchers that have used NESARC to find out if there is some simple explanation for our finding of an error in the coding of PG. We keep their identity anonymous, for reasons that will become apparent.

One responded:

Yes, NESARC is not easy to navigate. There are some sort of weights you need to add to the calculations, to weight by baseline characteristics because some subgroups were oversampled for the survey. I have not done this in many years, and no longer have the syntax. I believe they are described in many of the NESARC papers, as everybody appears to use these weights. My guess is that is what is incorrect, rather than how you’re coding the DSM criteria, which is more straightforward. In any case, I’m certain the numbers reported in the 2005 paper are correct, because many subsequent papers have been written about the PGs in that dataset, and no one’s reported difficulty replicating the n’s or analyses. I am not a statistician, but there are a lot of statisticians who are experienced at NESARC analyses and could probably help.

In fact, we find NESARC relatively straightforward to navigate, and it is well documented apart from this apparent coding error. The use of survey weights, which we are aware of and use for appropriate tabulations (e.g., prevalence and odd ratio estimation), is not the issue since this refers to the raw sample count reported in published papers. The final point, that we should trust that everyone else that has used the data has checked this, is not germane to a scholarly evaluation of these data. There is a simple error or there is not.

A second researcher, a statistician and experienced at NESARC analysis, then responded:

I have read your attachment and have 2 initial comments that might help resolve this issue. First the skip item is only used during the interview to decide whether to ask additional pg questions and does not come into play with regard to creating diagnoses. Second we never use the ever columns to determine diagnoses. We use the 12 month and prior to the 12 month columns only and if positive for any of those 2 time frames we declare a lifetime diagnosis. If this does not resolve your issues please let me know so I can examine them more closely.

The first point, about the skip question, is irrelevant to the issue. The second point is worth checking, and would arise from some inconsistency between reports made by the same individual about the same NESARC criteria. In fact, however, if one reviews the earlier extract from the questions involved, it is apparent that consistency here is enforced if the interviewer followed the logic of the skip pattern correctly. There are 4 such inconsistencies. In 3 of these 4 there was a positive
responses to the direct lifetime question, but both the “last 12 months” and “before the last 12
months” are left blank; perhaps the respondent said that they could not be bothered with the follow-
up. In 1 of these 4 the “last 12 months” was left blank but the “before 12 months” was marked NO.
This cannot be the cause of the discrepancy we are pointing out.

question is automatically filled in as a YES to ensure consistency.
Appendix B: Additional Documentation of Results (NOT FOR PUBLICATION)

All variables are defined from the original data files of wave 1 of the NESARC, and Stata code explaining the detailed definitions is available on request. The variables are as follows:

<table>
<thead>
<tr>
<th>variable name</th>
<th>variable label</th>
</tr>
</thead>
<tbody>
<tr>
<td>pg DSM_correct</td>
<td>Correct DSM-IV criteria applied to NESARC data for lifetime</td>
</tr>
<tr>
<td>alcohol</td>
<td>Any lifetime alcohol disorder</td>
</tr>
<tr>
<td>drug</td>
<td>Any lifetime drug use disorder</td>
</tr>
<tr>
<td>nicotine</td>
<td>Any lifetime nicotine disorder</td>
</tr>
<tr>
<td>mood</td>
<td>Any mood disorder</td>
</tr>
<tr>
<td>anxiety</td>
<td>Any anxiety disorder</td>
</tr>
<tr>
<td>personality</td>
<td>Any personality disorder</td>
</tr>
<tr>
<td>female</td>
<td>Female</td>
</tr>
<tr>
<td>black</td>
<td>Black or Afro-American</td>
</tr>
<tr>
<td>hispanic</td>
<td>Hispanic</td>
</tr>
<tr>
<td>age30_44</td>
<td>Aged between 30 and 44</td>
</tr>
<tr>
<td>age45_64</td>
<td>Aged between 45 and 64</td>
</tr>
<tr>
<td>age65plus</td>
<td>Aged 65 and over</td>
</tr>
<tr>
<td>separated</td>
<td>Separated by being widowed, divorced or separated</td>
</tr>
<tr>
<td>midwest</td>
<td>Midwest region</td>
</tr>
<tr>
<td>south</td>
<td>South region</td>
</tr>
<tr>
<td>west</td>
<td>West region</td>
</tr>
<tr>
<td>high school</td>
<td>Completed High School or GED</td>
</tr>
<tr>
<td>some college</td>
<td>Completed some college</td>
</tr>
<tr>
<td>college</td>
<td>Completed a college degree</td>
</tr>
<tr>
<td>graduate</td>
<td>Completed a graduate degree</td>
</tr>
<tr>
<td>income2</td>
<td>Personal income between $20,000 and $35,000</td>
</tr>
<tr>
<td>income3</td>
<td>Personal income between $35,000 and $70,000</td>
</tr>
<tr>
<td>income4</td>
<td>Personal income of $70,000 or more</td>
</tr>
<tr>
<td>height</td>
<td>Height in feet</td>
</tr>
<tr>
<td>weigh</td>
<td>Weight in stones</td>
</tr>
<tr>
<td>friday</td>
<td>Interview conducted on a Friday</td>
</tr>
<tr>
<td>saturday</td>
<td>Interview conducted on a Saturday</td>
</tr>
<tr>
<td>sunday</td>
<td>Interview conducted on a Sunday</td>
</tr>
<tr>
<td>pg Ntriggers</td>
<td>Number of triggers activated when pg trigger question asked</td>
</tr>
<tr>
<td>personality battery1</td>
<td>Responses to questions on usual feelings and actions</td>
</tr>
<tr>
<td>personality battery2</td>
<td>Responses to questions on effects of feelings and actions</td>
</tr>
<tr>
<td>ss 1</td>
<td>Sample Selection variable 1 from Screening Question #18 Part 1</td>
</tr>
<tr>
<td>ss 2</td>
<td>Sample Selection variable 2 from Screening Question #18 Part 2</td>
</tr>
<tr>
<td>ss 3</td>
<td>Sample Selection variable 3 from Screening Question #19 Part 1</td>
</tr>
<tr>
<td>ss 4</td>
<td>Sample Selection variable 4 from Screening Question #19 Part 2</td>
</tr>
<tr>
<td>ss 5</td>
<td>Sample Selection variable 5 from Screening Question #20</td>
</tr>
<tr>
<td>ss 6</td>
<td>Sample Selection variable 6 from Screening Question #21 Part 1</td>
</tr>
<tr>
<td>ss 7</td>
<td>Sample Selection variable 7 from Screening Question #21 Part 2</td>
</tr>
<tr>
<td>ss 8</td>
<td>Sample Selection variable 8 from Screening Question #21 Part 3</td>
</tr>
<tr>
<td>ss 9</td>
<td>Sample Selection variable 9 from Screening Question #22</td>
</tr>
<tr>
<td>ss 10</td>
<td>Sample Selection variable 10 from Screening Question #23 Part 1</td>
</tr>
<tr>
<td>ss 11</td>
<td>Sample Selection variable 11 from Screening Question #23 Part 2</td>
</tr>
<tr>
<td>ss 12</td>
<td>Sample Selection variable 12 from Screening Question #23 Part 3</td>
</tr>
<tr>
<td>ss 13</td>
<td>Sample Selection variable 13 from Screening Question #23 Part 4</td>
</tr>
<tr>
<td>ss 14</td>
<td>Sample Selection variable 14 from Screening Question #23 Part 5</td>
</tr>
<tr>
<td>ss 15</td>
<td>Sample Selection variable 15 from Screening Question #23 Part 6</td>
</tr>
<tr>
<td>ss 16</td>
<td>Sample Selection variable 16 from Screening Question #23 Part 7</td>
</tr>
<tr>
<td>ss 17</td>
<td>Sample Selection variable 17 from Screening Question #23 Part 8</td>
</tr>
<tr>
<td>ss 18</td>
<td>Sample Selection variable 18 from Screening Question #23 Part 9</td>
</tr>
</tbody>
</table>
Summary statistics are as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>pg_dsm_cor-t</td>
<td>43,093</td>
<td>.0046875</td>
<td>.0683057</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>alcohol</td>
<td>43,093</td>
<td>.2748242</td>
<td>.4464308</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>drug</td>
<td>43,093</td>
<td>.0944005</td>
<td>.2923885</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>nicotine</td>
<td>43,093</td>
<td>.1609774</td>
<td>.3675144</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>mood</td>
<td>43,093</td>
<td>.2111944</td>
<td>.4081607</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>anxiety</td>
<td>43,093</td>
<td>.1732764</td>
<td>.3784904</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>personality</td>
<td>43,093</td>
<td>.1460794</td>
<td>.3531899</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>female</td>
<td>43,093</td>
<td>.5702782</td>
<td>.4950421</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>black</td>
<td>43,093</td>
<td>.1995684</td>
<td>.3996806</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>hispanic</td>
<td>43,093</td>
<td>.1927923</td>
<td>.394496</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>age30_44</td>
<td>43,093</td>
<td>.3105377</td>
<td>.4627191</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>age45_64</td>
<td>43,093</td>
<td>.2979602</td>
<td>.4573672</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>age65plus</td>
<td>43,093</td>
<td>.1904022</td>
<td>.3926229</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>separated</td>
<td>43,093</td>
<td>.2579769</td>
<td>.4375263</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>midwest</td>
<td>43,093</td>
<td>.2086418</td>
<td>.4063425</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>south</td>
<td>43,093</td>
<td>.3749101</td>
<td>.4841053</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>west</td>
<td>43,093</td>
<td>.2559532</td>
<td>.4182133</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>high_school</td>
<td>43,093</td>
<td>.291161</td>
<td>.4543028</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>some_college</td>
<td>43,093</td>
<td>.2063212</td>
<td>.4046685</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>college</td>
<td>43,093</td>
<td>.2093844</td>
<td>.4068739</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>graduate</td>
<td>43,093</td>
<td>.1109925</td>
<td>.3141265</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>income2</td>
<td>43,093</td>
<td>.2314761</td>
<td>.4217808</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>income3</td>
<td>43,093</td>
<td>.2095236</td>
<td>.4063733</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>income4</td>
<td>43,093</td>
<td>.2906969</td>
<td>.4540929</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>height</td>
<td>43,093</td>
<td>5.549293</td>
<td>.3356052</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>weigh</td>
<td>43,093</td>
<td>12.17406</td>
<td>2.914927</td>
<td>4.428571</td>
<td>35.71429</td>
</tr>
<tr>
<td>friday</td>
<td>43,093</td>
<td>.1413919</td>
<td>.3484294</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>saturday</td>
<td>43,093</td>
<td>.1438981</td>
<td>.3509904</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>sunday</td>
<td>43,093</td>
<td>.0768106</td>
<td>.2662938</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>pg_Ntriggers</td>
<td>43,093</td>
<td>.1886385</td>
<td>.7641052</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>personalit-1</td>
<td>43,093</td>
<td>3.466642</td>
<td>4.772873</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>personalit-2</td>
<td>43,093</td>
<td>3.443552</td>
<td>4.768956</td>
<td>0</td>
<td>55</td>
</tr>
</tbody>
</table>

ss_1          | 43,093 | .348154  | .4763907 | 0   | 1   |
ss_2          | 43,093 | .3080841 | .4586477 | 0   | 1   |
ss_3          | 43,093 | .2764718 | .4472581 | 0   | 1   |
ss_4          | 43,093 | .2422435 | .4284459 | 0   | 1   |
ss_5          | 43,093 | .3640963 | .4811814 | 0   | 1   |
ss_6          | 43,093 | .76256   | .425519  | 0   | 1   |
ss_7          | 43,093 | .7988697 | .4007753 | 0   | 1   |
ss_8          | 43,093 | .4865291 | .4998243 | 0   | 1   |
ss_9          | 43,093 | .2665166 | .4421426 | 0   | 1   |
ss_10         | 43,093 | .3197503 | .4663851 | 0   | 1   |
ss_11         | 43,093 | .3476435 | .4762276 | 0   | 1   |
ss_12         | 43,093 | .1458938 | .3530038 | 0   | 1   |
ss_13         | 43,093 | .0624463 | .2419673 | 0   | 1   |
ss_14         | 43,093 | .0896897 | .2857401 | 0   | 1   |
ss_15         | 43,093 | .0792224 | .2700912 | 0   | 1   |
ss_16         | 43,093 | .2083633 | .4061427 | 0   | 1   |
ss_17         | 43,093 | .0647437 | .2460759 | 0   | 1   |
The sample selection screening questions come from the “Background Information” section of the NESARC questionnaire, reproduced below.

### 18. During the flashcard 11C

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

### 19. During the flashcard 11C

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.
2. Didn’t do work or other activities as carefully as usual.

### 20. During the flashcard 11D

During the past 4 weeks, how much did pain interfere with your normal work including both work outside the home and housework?
(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?
   - All of the time
   - Most of the time
   - Some of the time
   - A little of the time
   - None of the time

2. Did you have a lot of energy?
   - All of the time
   - Most of the time
   - Some of the time
   - A little of the time
   - None of the time

3. Have you felt downhearted and depressed?
   - All of the time
   - Most of the time
   - Some of the time
   - A little of the time
   - 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?
   - All of the time
   - Most of the time
   - Some of the time
   - A little of the time
   - None of the time
Figures B1, B2 and B3 show distributions of the predicted marginal probabilities of each gambling disorder category for the DSM-IV screen from the *British Gambling Prevalence Survey* of 2010, the PGSI screen of the *Victorian Gambling Survey* of 2008, and the NODS screen of the Victorian Gambling Survey of 2008. In some cases the horizontal axes include labels for probabilities below 0 or above 1: this is an artefact of the kernel density procedure used to display distributions as well as the use of the “delta method” to approximate non-linear combinations of estimated parameters. These are of no substantive or statistical significance for our purposes.
Figure B1: Statistical Significance of Sample Selection Corrections for Gambling Risk in the U.K., Measured with the DSM-IV Screen

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

Source: British Gambling Prevalence Survey of 2010 and author estimates

Non-Indicated

Uncorrected
Corrected

90th percentile p-value = 0.804
99th percentile p-value = 0.901

Weakly Indicated

Uncorrected
Corrected

90th percentile p-value = 0.381
99th percentile p-value = 0.651
99th percentile p-value = 0.927

Moderately Indicated

Uncorrected
Corrected

90th percentile p-value = 0.752
99th percentile p-value = 0.871
99th percentile p-value = 0.975

Pathological Gambler

Uncorrected
Corrected

90th percentile p-value = 0.126
99th percentile p-value = 0.377
99th percentile p-value = 0.856
Figure B2: Statistical Significance of Sample Selection Corrections for Gambling Risk in Victoria (Australia), Measured with the PGSI Screen

100 predicted marginal probabilities from each model, for each individual, reflecting covariance of estimates
Source: Wave 1 of the *Victorian Gambling Survey* of 2008 and author estimates

Non-Indicated

Uncorrected
Corrected

90th percentile p-value = 0.010
95th percentile p-value = 0.177
99th percentile p-value = 0.803

Weakly Indicated

Uncorrected
Corrected

90th percentile p-value = 0.001
95th percentile p-value = 0.069
99th percentile p-value = 0.724

Moderately Indicated

Uncorrected
Corrected

90th percentile p-value = 0.394
95th percentile p-value = 0.664
99th percentile p-value = 0.952

Pathological Gambler

Uncorrected
Corrected

90th percentile p-value = 0.362
95th percentile p-value = 0.656
99th percentile p-value = 0.925
Figure B3: Statistical Significance of Sample Selection Corrections for Gambling Risk in Victoria (Australia), Measured with the NODS Screen

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

Source: Wave 1 of the Victorian Gambling Survey of 2008 and author estimates

Non-Indicated

Weakly Indicated

Moderately Indicated

Pathological Gambler
Appendix C: Formal Statistical Model (NOT FOR PUBLICATION)

The statistical model is an extension of the standard Heckman model in two 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.

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

\[ Y^*_j = \beta_j^T X_j + U_j \quad j = 1, 2 \quad (1) \]

\[ Y_1 = I( Y_1^* \geq 0 ) \quad (2) \]

\[ Y_2 = \sum_{h=0}^{H} h I( \alpha_h < Y_2^* \leq \alpha_{h+1} ) \quad \text{if } Y_1 = 1 \quad (3) \]

where \( Y_1^* \) is a continuous latent variable for the sample selection equation, \( Y_2^* \) is a continuous latent variable for the risk of GD, \( \beta_j \) denotes \( k_j \) vectors of parameters to be estimated, \( X_j \) denotes \( k_j \) vectors of exogenous variables, the \( U_j \) are random errors, \( I(\cdot) \) 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 GD, \( H+1 \) denotes the ordered categories of GD, and \( (\alpha_0, ..., \alpha_h, \alpha_{h+1}, ..., \alpha_H) \) 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 GD. The \( H+1 \) cutpoints \( (\alpha_0, ..., \alpha_h, \alpha_{h+1}, ..., \alpha_H) \), 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 GD.\(^{43}\)

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 \( \beta_2 \) 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

\(^{43}\) 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 \( \rho \). 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.
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. 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 $\beta_1$ and the first threshold $\alpha_1$ 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]. The software runs on Stata version 15, although all estimations run on Stata version 14.

Additional References


