Journal of Gambling Studies (2020) 36:1133–1159
https://doi.org/10.1007/s10899-019-09897-2

ORIGINAL PAPER

The Risk of Gambling Problems in the General Population: A Reconsideration

Glenn W. Harrison[1,4,5] · **Morten I. Lau**[2,6] · **Don Ross**[3,4,5]

Published online: 18 October 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
We examine the manner in which the population prevalence of disordered gambling has usually been estimated, on the basis of surveys that suffer from a potential sample selection bias. General population surveys screen respondents using seemingly innocuous "trigger," "gateway" or "diagnostic stem" questions, applied before they ask the actual questions about gambling behavior and attitudes. Modeling the latent sample selection behavior generated by these trigger questions using up-to-date econometrics for sample selection bias correction 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 standard mental health surveys, and not just gambling disorder. We discuss ways in which these behavioral biases can be mitigated in future surveys.

Keywords Gambling disorder · Prevalence studies · Sample selection bias · Bias correction · Econometrics · Diagnostic stem questions · Comorbidities

Prevalence studies of disordered gambling have been conducted in many countries over three decades (Williams et al. 2012). In consequence there is widespread consensus that gambling disorder, at least of a level of severity warranting clinical intervention, is a relatively rare mental disorder, though one that has become more common in many jurisdictions as a result of more widespread gambling opportunities. Based on the application of econometric methods for identification and control of sample selection bias, we question this consensus, concluding that prevalence of gambling problems in the general population is likely to be significantly larger than generally thought. The issues identified here apply,

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10899-019-09897-2) contains supplementary material, which is available to authorized users.

✉ Don Ross
don.ross931@gmail.com; don.ross@uct.ac.za

Extended author information available on the last page of the article

in principle, to every disorder for which prevalence is estimated using surveys based on psychiatric screening instruments, and not just gambling disorder.

Scholarly research consistently finds high shares of commercial gambling revenue to be derived from proportions of populations that are much smaller than 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.[1] 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. 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).

We 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.[2] 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 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 than are typically reported. 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 prevalence 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. Since the classification of the condition under "Substance-Related and Addictive Disorders" in *DSM 5* (American Psychiatric Association 2013), it has become standard usage to refer to it as "Gambling Disorder".

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 some work applying the Problem Gambling Severity Index (PGSI) (Ferris and Wynne 2001), we will anachronistically use the terms "gambling disorder" and "disordered gambler." Where we

[1] See Gerstein 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.

[2] Kessler and Pennell (2015, p. 144ff.) provide a valuable review of the historical evolution of survey research on mental disorders.

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. 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.[3] 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 *DSM 5* labels as GD.[4]

In Sect. 1 we reconsider inferences from the *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.[5] The instrument for measuring gambling problems was based on the *DSM-IV* criteria, and *DSM-IV* criteria were likewise used for the instruments measuring other major psychiatric disorders.[6]

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 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

[3] Examples of the many studies of the continuum of gambling disorders include Toce-Gerstein et al. (2003) and Blanco et al. (2006).

[4] 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 the choice 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.

[5] 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 (NIAAA) and does not reflect the opinions or views of NIAAA or the U.S. Government.

[6] A comparable national survey that could also be evaluated in the same manner is the National *Comorbidity Survey Replication* conducted in the United States between 2001 and 2003 with a primary sample of 9282 individuals. We discuss the *Canadian Community Health Survey of Mental Health and Well-Being* of 2002 and the *British Gambling Prevalence Survey* of 2010 below.

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.[7] Sometimes this potential leads to no difference in inferences from the observed sample: for instance, if respondents want to spend more time in a face-to-face interview with more attractive interviewers, and the attractiveness level of interviewers is random, there will be no a priori reason to expect an effect on inferences about 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 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 Sect. 2 to review the rationale for these trigger questions, and note the vigorous 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.

Estimates for the United States from NESARC

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

[7] Hernán et al. (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 et al. (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 et al. 1996; 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.".

Table 1 Logistic odds ratio calculations for pathological gambling

	Odds ratio	Standard error	t for H_0: OR = 1	p value	95% confidence interval	
A. *Total effects with logistic model and no covariates*						
Alcohol	6.1	0.52	21.1	< 0.001	5.1	7.2
Drug	5.2	0.61	14.3	< 0.001	4.2	6.6
Nicotine	7.0	0.62	22.0	< 0.001	5.8	8.3
Mood	4.8	0.44	17.3	< 0.001	4.0	5.8
Anxiety	3.4	0.32	13.1	< 0.001	2.8	4.1
Personality	9.0	0.81	24.2	< 0.001	7.4	10.7
B. *Marginal effects with logistic model and no covariates*						
Alcohol	2.6	0.19	13.1	< 0.001	2.2	3.0
Drug	1.3	0.14	2.2	0.03	1.0	1.6
Nicotine	2.7	0.21	13.2	< 0.001	2.4	3.2
Mood	1.7	0.21	4.2	< 0.001	1.3	2.1
Anxiety	1.1	0.13	0.9	0.39	0.9	1.4
Personality	4.1	0.65	8.8	< 0.001	2.9	5.6
C. *Total effects with logistic model and covariates*						
Alcohol	5.2	0.48	17.8	< 0.001	4.3	6.2
Drug	3.9	0.45	11.9	< 0.001	3.1	4.9
Nicotine	6.1	0.51	21.4	< 0.001	5.1	7.2
Mood	5.1	0.52	16.3	< 0.001	4.2	6.3
Anxiety	3.8	0.35	14.5	< 0.001	3.2	4.6
Personality	7.8	0.68	23.6	< 0.001	6.6	9.3
D. *Marginal effects with logistic model and covariates*						
Alcohol	2.2	0.18	9.4	< 0.001	1.8	2.6
Drug	1.1	0.13	1.2	0.24	0.9	1.4
Nicotine	2.8	0.22	13.5	< 0.001	2.4	3.3
Mood	1.9	0.26	5.0	< 0.001	1.5	2.5
Anxiety	1.3	0.15	2.1	0.04	1.0	1.6
Personality	3.7	0.58	8.4	< 0.001	2.7	5.1
E. *Marginal odds ratios for demographic covariates with logistic model*						
Female	**0.4**	**0.04**	**− 9.3**	**< 0.001**	**0.3**	**0.5**
Black	**3.1**	**0.25**	**14.1**	**< 0.001**	**2.7**	**3.7**
Hispanic	0.9	0.11	− 1.2	0.24	0.7	1.1
Age 30–44	0.9	0.17	− 0.3	0.77	0.7	1.4
Age 45–64	1.3	0.22	1.7	0.09	0.9	1.8
Age 65+	1.1	0.18	0.7	0.51	0.8	1.5
Separated	**1.5**	**0.15**	**4.3**	**< 0.001**	**1.2**	**1.8**
Midwest	1.1	0.14	0.8	0.45	0.8	1.4
South	0.9	0.10	− 0.9	0.35	0.7	1.1
West	**1.7**	**0.22**	**4.0**	**< 0.001**	**1.3**	**2.2**
High school	0.9	0.13	− 0.1	0.91	0.7	1.3
Some college	1.0	0.08	− 0.3	0.78	0.8	1.1
College	**0.5**	**0.07**	**− 5.3**	**< 0.001**	**0.4**	**0.6**
Graduate	**0.7**	**0.13**	**− 2.2**	**0.03**	**0.4**	**0.9**

Table 1 (continued)

	Odds ratio	Standard error	t for H_0: OR = 1	p value	95% confidence interval	
Income 2	0.9	0.12	−0.4	0.72	0.7	1.2
Income 3	**0.6**	**0.06**	**−5.3**	**< 0.001**	**0.5**	**0.7**
Income 4	**1.5**	**0.17**	**3.3**	**0.002**	**1.2**	**1.9**

obtains the same results, using a logistic specification.[8] All calculations with the NESARC correct for the complex sampling design.[9] In each row the independent binary variable is whether the respondent is defined as having the indicated psychiatric disorder or not.[10] 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.[11] 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

[8] 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.

[9] 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*.

[10] 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.

[11] Epidemiologists often report "adjusted odds ratios," which control for covariates. Typically the list of covariates is very small.

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).[12] If this suggestion is correct, it has direct implications for 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.[13] 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.

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 $70 k at the time of the survey. The value of these OR estimates, and their 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

[12] 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 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.

[13] This is "model 3" of Petry et al. (2005).

indicators.[14] 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 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.[15] 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.[16]

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."

[14] 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.

[15] 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).

[16] 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.

Often this is a major empirical challenge, since it can be hard to exclude something from potentially affecting the main variable of interest, but to include it as likely to affect sample selection. In epidemiology, for instance, a spirited defence[17] 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 *particularly* "problematic" in the absence of 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.[18] 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 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 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 Fig. 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.

[17] Criticisms were raised by Geneletti et al. (2011) in response to epidemiological applications of corrections for sample selection by Chaix et al. (2011) and Bärnighausen et al. (2011a).

[18] 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.

Table 2 Semi-nonparametric odds ratio calculations for pathological gambling

	Odds ratio	Standard error	95% confidence interval	
A. *Total effects with no covariates*				
Alcohol	6.0	< 0.001	6.0	6.0
Drug	6.2	< 0.001	6.2	6.2
Nicotine	6.7	< 0.001	6.7	6.7
Mood	4.8	< 0.001	4.8	4.8
Anxiety	4.2	< 0.001	4.2	4.2
Personality	8.2	< 0.001	8.2	8.2
B. *Marginal effects with no covariates*				
Alcohol	4.2	0.004	4.2	4.3
Drug	1.6	0.001	1.6	1.6
Nicotine	3.3	0.003	3.3	3.3
Mood	2.4	0.002	2.4	2.4
Anxiety	2.3	0.001	2.3	2.3
Personality	4.6	0.005	4.6	4.7

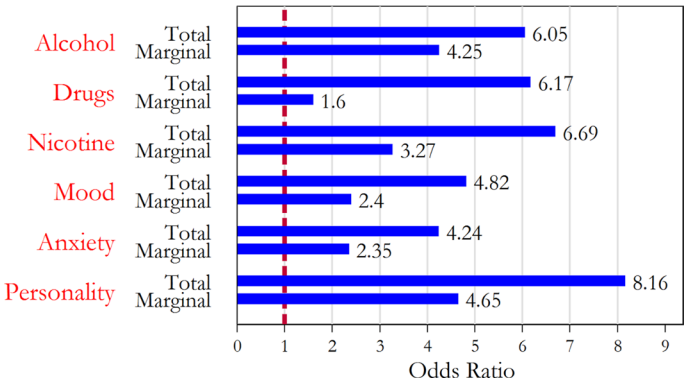

Fig. 1 Comorbidity of gambling disorder and other psychiatric disorders. Estimated odds ratios using semi-nonparametric ordered response model. *Source*: National Epidemiological Survey on Alcohol and Related Conditions (NESARC)

Figure 2 shows marginal effects of comorbidities when one undertakes sample selection corrections.[19] 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.[20] 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,

[19] 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.

[20] Appendix B documents these covariates.

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, 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.

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."[21] 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.[22] We retain 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%

[21] 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.

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

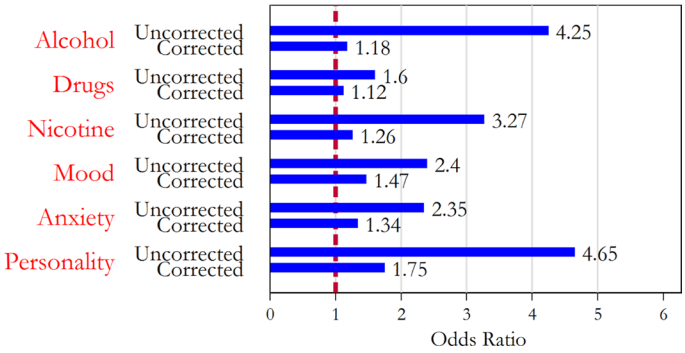

Fig. 2 Effect of sample selection correction on estimates of comorbidity marginal effects with gambling disorder. Estimated odds ratios using semi-nonparametric ordered response model. *Source*: National Epidemiological Survey on Alcohol and Related Conditions (NESARC)

Table 3 Semi-nonparametric odds ratio calculations for gambling risks

	Odds ratio with no sample selection correction	Odds ratio with sample selection correction
A. *Weakly indicated*		
Alcohol	1.95	1.02
Drug	1.25	1.02
Nicotine	1.72	1.03
Mood	1.50	1.05
Anxiety	1.48	1.04
Personality	2.01	1.07
B. *Moderately indicated*		
Alcohol	2.91	1.10
Drug	1.42	1.06
Nicotine	2.40	1.14
Mood	1.91	1.24
Anxiety	1.88	1.18
Personality	3.08	1.35
C. *Pathological gambler*		
Alcohol	4.25	1.18
Drug	1.60	1.12
Nicotine	3.27	1.26
Mood	2.40	1.47
Anxiety	2.35	1.34
Personality	4.65	1.75

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

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 Figs. 3 and 4. We therefore find a common result, that the prevalence of Pathological Gambling 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; 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 much of 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 1.3% when no correction is applied.

It is worth stressing that this result obtains 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%.[23] 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.[24]

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 Fig. 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 SNP ordered probit model. Each random draw is from a normal distribution whose mean is the point estimate of the

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

[24] 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 1754 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 is 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 for GD.

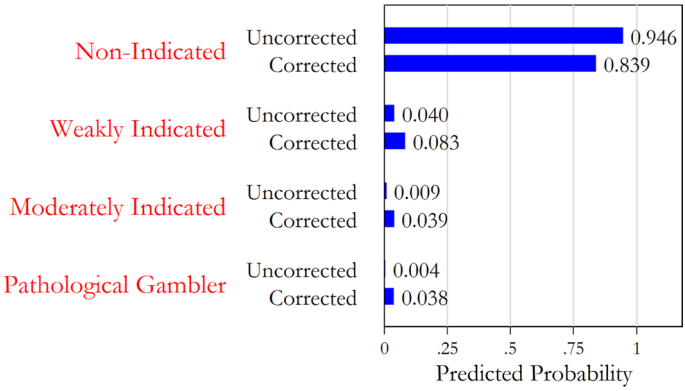

Fig. 3 Predicted prevalence of gambling risk with and without sample selection correction. Estimated probabilities using semi-nonparametric ordered response model. *Source*: National Epidemiological Survey on Alcohol and Related Conditions (NESARC)

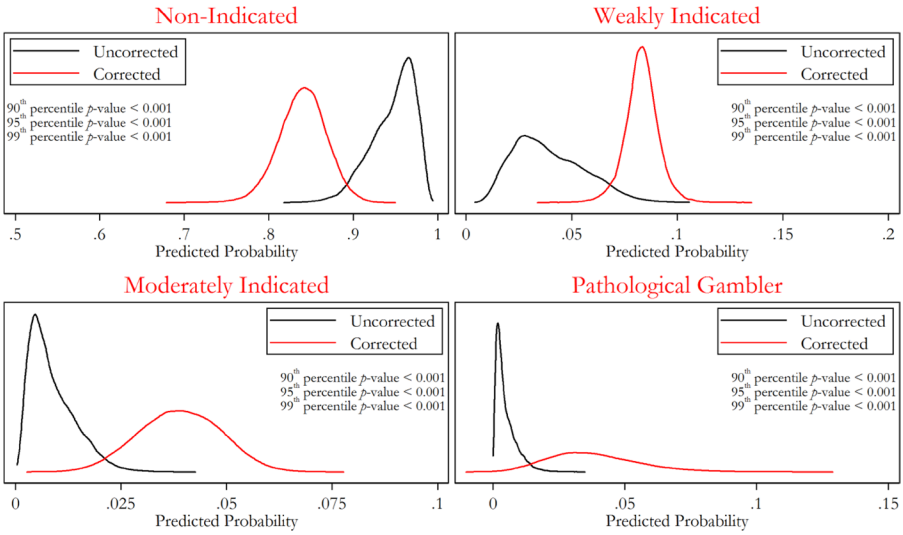

Fig. 4 Statistical significance of sample selection corrections for gambling risk. 100 predicted marginal probabilities from each model, for each individual, reflecting covariance of estimates

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 Fig. 4 reflect 4,309,300 predictions.

These densities in Fig. 4 allow one to see the average effects shown in Fig. 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 Fig. 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 0.001 or lower.[25] Similarly, the average predicted *increases* in the Weakly Indicated, Moderately Indicated and Pathological Gambler categories (Fig. 3) are also statistically significant, with the 99th percentile of *p* values again being 0.001 or lower in each case (Fig. 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 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 Fig. 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.

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.

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 dependency, and was prompted by knowledge from previous clinical treatment of the correlations

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

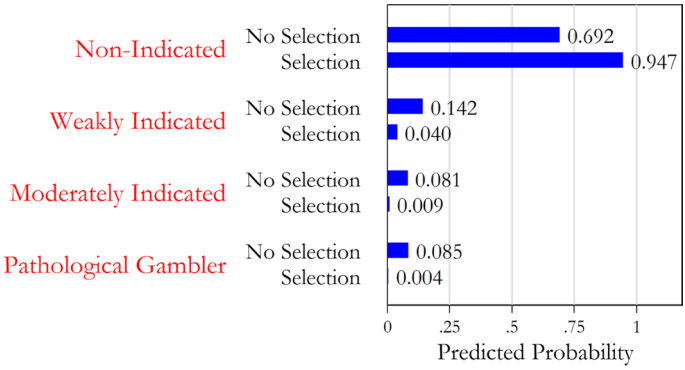

Fig. 5 Predicted probability of gambling risk conditional on sample selection or not. Predicted probability of sample selection = 0.58. Estimated probabilities using semi-nonparametric ordered response model. *Source*: National Epidemiological Survey on Alcohol and Related Conditions (NESAKC)

between these addictions and gambling problems (Lesieur et al. 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.[26]

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 et al. (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 public health forecasting.

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

[26] 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.

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 et al. (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 et al. (2012) shows. There are continuing debates about the nature of those trigger questions, but they generally concern whether participants should be asked about 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 contribute 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[27] 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 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 et al. (1977) 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. (1977; p. 249ff.) suggest

[27] The citation, strictly speaking, refers to the PGSI, the short scored field screen of the CPGI.

> 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.[28]

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 identified here (see Harrison et al. 2018). 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.

[28] 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.

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 7756: 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.[29] 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).[30] Again, Fig. 7 shows that although there are economically significant differences in prevalence overall when correcting for sample selection bias using the PGSI, the differences for the most important categories of Moderately Indicated and Pathological Gambler[31] are not statistically significant[32] 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.[33] Starting with the 75% of the sample that had gambled at all in the last 12 months, these surveyors added 1057 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

[29] 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 insignificance of the differences.

[30] 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).

[31] For consistency we repeat the categories of gambling problems and risk used in the NESARC data analysis (Figs. 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 Fig. 6 the original category is "Problem Gambling," with a *DSM-IV* score of 3 or more. In Fig. 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 Fig. 8 the original categories are "At Risk," "Problem Gambler," and "Pathological Gambler," respectively.

[32] Figure B2 in Appendix B documents the claim about statistical insignificance of the differences.

[33] 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 1759 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).

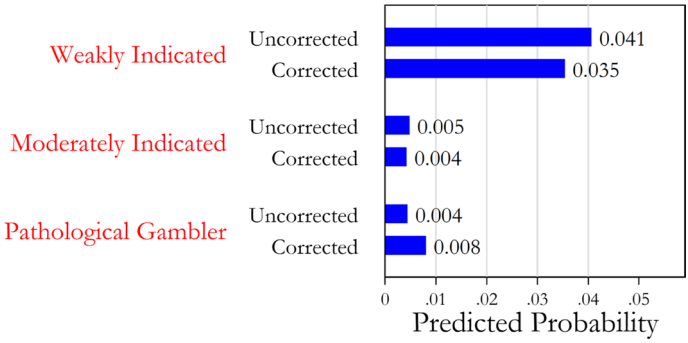

Fig. 6 Predicted prevalence of gambling disorders in the U.K., measured with the DSM-IV screen, with and without sample selection correction. Estimated probabilities using semi-nonparametric ordered response model. *Source*: British Gambling Prevalence Survey of 2010. Fraction at any risk level changes from 0.050 to 0.048 with correction

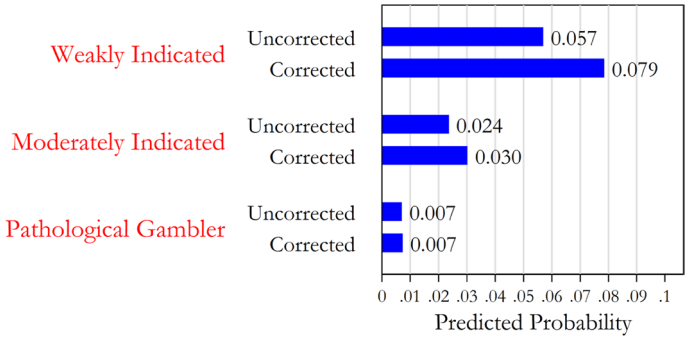

Fig. 7 Predicted prevalence of gambling disorders in Victoria (Australia), measured with the PGSI screen, with and without sample selection correction. Estimated probabilities using semi-nonparametric ordered response model. *Source*: Wave 1 of the Victorian Gambling Survey of 2008. Fraction at any risk level changes from 0.088 to 0.116 with correction.

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 Fig. 8 shows, it leads to a statistically and quantitatively significant sample selection bias when inferring lifetime prevalence.[34] 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

[34] Figure B3 in Appendix B documents the claim about statistical significance of the differences.

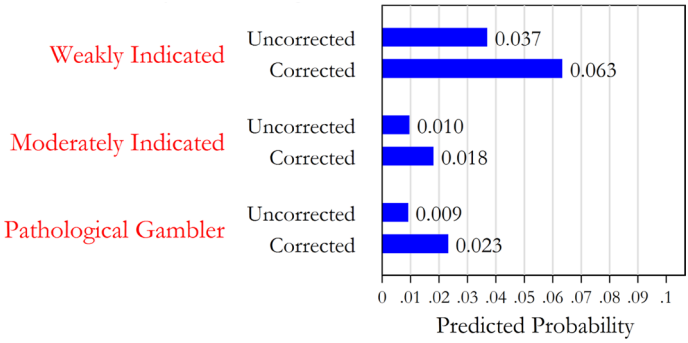

Fig. 8 Predicted prevalence of gambling disorders in Victoria (Australia), measured with the NODS screen, with and without sample selection correction. Estimated probabilities using semi-nonparametric ordered response model. *Source*: Wave 1 of the Victorian Gambling Survey of 2008. Fraction at any risk level changes from 0.056 to 0.105 with correction

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. (2015a, b) 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 et al. 2018) 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.[35] 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 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.[36] An example of the latter would be any one of a myriad of survey techniques for using "randomized response"

[35] 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.

[36] 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 et al. (2006). This randomized follow-up was in addi-

methods to ensure that subjects are not revealing with certainty some sensitive information in response to a question.[37]

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.

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 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.[38] Although it is conceivable that the distortion could be in any direction, our a priori expectation would

Footnote 36 (continued)

tion to the primary randomization to treatment: (1) a housing voucher with some strings attached and some counseling, (2) a housing voucher with no strings attached and no counseling, and (3) 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 et al. (2014), where subjects were offered different non-risky incentives to participate and effects on measured risk aversion assessed after correcting for sample selection.

[37] See Warner (1965) for the original idea, and Blaire et al. (2015) for a recent review and application of variants. There are other non-randomized survey methods for encouraging truthful responses to sensitive questions.

[38] For example, Harrison (2017) applies the same approach to the population prevalence of nicotine dependence, which is DSM-IV code 305.10, and finds comparable biases in the United States using NESARC.

be that the distortions lead to understatements of prevalence across the board, given the sensitive nature of the trigger questions.

Funding 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.

Compliance with Ethical Standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical Standards All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

Abbott, M. W., & Volberg, R. A. (2000). *Taking the pulse on gambling and problem gambling in New Zealand: A report on phase one of the 1999 national prevalence survey*. New Zealand: Department of Internal Affairs, Government of New Zealand.

Algeria, A. A., Petry, N. M., Hasin, D. S., Liu, S.-M., Grant, B. F., & Blanco, C. (2009). Disordered gambling among racial and ethnic groups in the US: Results from the national epidemiologic survey on alcohol and related conditions. *CNS Spectrums, 14,* 132–142.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders III revision (DSM-III-R)*. Washington, DC: APA Press.

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

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

Australian Productivity Commission. (1999). *Australia's gambling industries: Inquiry report*. Canberra: Australian Government Productivity Commission.

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

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

Billi, R., Stone, C. A., Abbott, M., & Yeung, K. (2015). The Victorian Gambling Study (VGS): A longitudinal study of gambling and health in Victoria, 2008–2012: Design and methods. *International Journal of Mental Health and Addiction, 13,* 274–296.

Billi, R., Stone, C. A., Marden, P., & Yeung, K. (2014). *The Victorian gambling study: A longitudinal study of gambling and health in Victoria, 2008–2012*. North Melbourne: Victorian Responsible Gambling Foundation.

Blaire, G., Imai, K., & Zhou, Y.-Y. (2015). Design and Analysis of the randomized response technique. *Journal of the American Statistical Association, 110*(511), 1304–1319.

Blanco, C., Hasin, D. S., Petry, N., Stinson, F. S., & Grant, B. F. (2006). Sex differences in subclinical and DSM-IV pathological gambling: Results from the national epidemiologic survey on alcohol and related conditions. *Psychological Medicine, 36,* 943–953.

Blaszczynski, A., Dumlao, V., & Lange, M. (1977). 'How much do you spend on gambling?' Ambiguities in survey questionnaire form. *Journal of Gambling Studies, 13*(3), 237–252.

Caetano, R. (2001). Non-response in alcohol and drug surveys: A research topic in need of further attention. *Addiction, 96,* 1541–1545.

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

Currie, S. R., Miller, N., Hodgins, D. C., & Wang, J. L. (2009). Defining a threshold of harm from gambling for population health surveillance research. *International Gambling Studies, 9*(1), 19–38.

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

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

Dickerson, M. G., Baron, E., Hong, S.-M., & Cottrell, D. (1996). Estimating the extent and degree of Gambling related problems in the Australian population: A national survey. *Journal of Gambling Studies, 12*, 161–178.

DiNardo, J., McCrary, J., & Sanbonmatsu, L. (2006). *Constructive proposals for dealing with attrition: An empirical example*. NBER working paper.

Ferris, J., & Wynne, H. (2001). *The Canadian Problem Gambling Index final report*. Ottawa: Canadian Center on Substance Abuse. www.ccsa.ca/pdf/ccsa-008805-2001.pdf.

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

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

Gerstein, D., Hoffman, J., Larison, C., Engelman, L., Murphy, S., Palmer, A., et al. (1999). *Gambling impact and behavior study: Report to the National Gambling Impact Study Commission*. Chicago: National Opinion Research Center at the University of Chicago.

Harrison, G. W. (2017). Behavioral responses to surveys about nicotine dependence. *Health Economics, 26*, 114–123.

Harrison, G. W., Il, H., & Lau, M. (2014). *Risk attitudes, sample selection and attrition in a longitudinal field experiment*. CEAR working paper 2014-04. Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University. *Review of Economics and Statistics* (forthcoming).

Harrison, G. W., Jessen, L. J., Lau, M., & Ross, D. (2018). Disordered gambling prevalence: Methodological innovations in a general Danish population survey. *Journal of Gambling Studies, 34,* 225–253.

Harrison, G. W., & Ng, J. M. (2016). Evaluating the expected welfare gain from insurance. *Journal of Risk and Insurance, 83*(1), 91–120.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement, 5,* 475–492.

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

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

Kessler, R. C., & Pennell, B.-E. (2015). Developing and selecting mental health measures. In T. P. Johnson (Ed.), *Handbook of health survey methods*. New York: Wiley.

Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica, 51,* 507–512.

Lesieur, H. R. (1994). Epidemiological surveys of pathological gambling: Critique and suggestions for modification. *Journal of Gambling Studies, 10*(4), 385–398.

Lesieur, H. R., & Blume, S. B. (1987). The South Oaks Gambling Screen (SOGS): A new instrument for the identification of pathological gamblers. *American Journal of Psychiatry, 144*(9), 1184–1188.

Lesieur, H. R., Blume, S. B., & Zoppa, R. M. (1986). Alcoholism, drug abuse, and gambling. *Alcoholism, Clinical and Experimental Research, 10*(1), 33–38.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. New York: Cambridge University Press.

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

Nower, L., Martins, S., Lin, K.-H., & Blanco, C. (2013). Subtypes of disordered gamblers: Results from the national epidemiologic survey on alcohol and related conditions. *Addiction, 108*(4), 789–798.

Orr, L., Feins, J. D., Jacob, R., Beecroft, E., Sanbonmatsu, L., Katz, L. F., et al. (2003). *Moving to opportunity interim impacts evaluation*. Final Report. U.S. Department of Housing and Urban Development, 2003.

Petry, N., Stinson, F. S., & Grant, B. F. (2005). Comorbidity of DSM-IV oathological gambling and other psychiatric disorders: Results from the national epidemiological survey on alcohol and related conditions. *Journal of Clinical Psychiatry, 66*(5), 564–674.

Pietrzak, R. H., Morasco, B. J., Blanco, C., Grant, B. F., & Petry, N. M. (2007). 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,* 301–313.

Rachlin, H. (1990). Why do people gamble and keep gambling despite heavy losses? *Psychological Science, 1,* 294–297.

Rachlin, H. (2000). *The science of self-control*. Cambridge, MA: Harvard University Press.

Rachlin, H., Safin, V., Arfer, K. B., & Yen, M. (2015). The attraction of gambling. *Journal of the Experimental Analysis of Behavior, 103*(1), 260–266.

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

Schellinck, T., Schrans, T., Bliemel, M., & Schellinck, H. (2015a). Construct development for the focal adult gambling screen (FLAGS): A risk measurement for gambling harm and problem gambling associated with electronic gambling machines. *Journal of Gambling Issues, 30,* 140–173.

Schellinck, T., Schrans, T., Bliemel, M., & Schellinck, H. (2015b). Instrument development for the focal adult gambling screen (FLAGS-EGM): A measurement of risk and problem gambling associated with electronic gambling machines. *Journal of Gambling Issues, 30,* 174–200.

Schüll, N. D. (2012). *Addiction by design: Machine gambling in Las Vegas*. Princeton: Princeton University Press.

Sharp, C., Steinberg, L., Yaroslavsky, I., Hofmeyr, A., Dellis, A., Ross, D., et al. (2012). An item response theory analysis of the Problem Gambling Severity Index. *Assessment, 19*(2), 167–175.

StataCorp. (2013). *Stata base reference manual: Version 13*. College Station, TX: StataCorp LP.

Statistics Canada, Health Statistics Division. (2004). *Canadian Community Health Survey. Cycle 1.2: Mental health and well-being*. Ottawa: Statistics Canada, Catalogue #82M0021GPE.

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

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

Stone, C., Romild, U., Abbott, M., Young, K., Billi, R., & Volberg, R. (2015). Effects of different screening and scoring thresholds on PGSI gambling risk segments. *International Journal of Mental Health and Addiction, 13,* 82–102.

Tam, T. W., & Midanik, L. T. (2000). The effect of screening on prevalence estimates of alcohol dependence and social consequences. *Journal of Studies on Alcohol, 61*(4), 617–621.

Tam, T. W., Midanik, L. T., Greenfield, T. K., & Caetano, R. (1996). Selection bias in national surveys due to screening: implications from a county general population survey. *Addiction, 91*(4), 557–564.

Toce-Gerstein, M., Gerstein, D., & Volberg, R. (2003). A hierarchy of gambling disorders in the community. *Addiction, 98,* 1661–1672.

Van de Ven, W. P. M. M., & Van Praag, B. M. S. (1981). The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of Econometrics, 17,* 229–252.

Volberg, R. A. (1996). Prevalence studies of problem gambling in the United States. *Journal of Gambling Studies, 12*(2), 111–128.

Volberg, R. A., & Steadman, H. J. (1988). Refining prevalence estimates of pathological gambling. *American Journal of Psychiatry, 145*(4), 502–505.

Volberg, R. A., & Steadman, H. J. (1989). Prevalence estimates of pathological gambling in New Jersey and Maryland. *American Journal of Psychiatry, 146*(12), 1618–1619.

Volberg, R. A., & Williams, R. J. (2012). *Developing a short form of the PGSI*. Report to the Gambling Commission. Northampton, MA: Gemini Research. http://www.gamblingcommission.gov.uk/pdf.

Wardle, H., Moody, A., Spence, S., Orford, J., Volberg, R., Jotangia, D., et al. (2011). *British Gambling Prevalence Survey 2010*. London: National Centre for Social Research.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60*(309), 63–69.

Williams, R. J., & Volberg, R. (2009). Impact of survey description, administration format, and exclusionary criteria on population prevalence rates of problem gambling. *International Gambling Studies, 9*(2), 101–117.

Williams, R. J., & Volberg, R. A. (2010). *Best practices in the population assessment of problem gambling*. Report. Guelph: Ontario Problem Gambling Research Centre. https://www.uleth.ca/dspace/handle/10133/1259.

Williams, R. J., Volberg, R. A., & Stevens, R. M. G. (2012). *The population prevalence of problem gambling: Methodological influences, standardized rates, jurisdictional differences, and worldwide trends*. Report Prepared for the Ontario Problem Gambling Research Centre & the Ontario Ministry of Health and Long Term Care. www.gamblingresearch.org/applydownload.php?docid=12579?.

Williams, R. J., & Wood, R. T. (2004). The proportion of gambling revenue derived from problem gamblers: Examining the issues in a Canadian context. *Analyses of Social Issues and Public Policy, 4*(1), 33–45.
Wood, R. T., & Williams, R. J. (2007). 'How much money do you spend on gambling?' The comparative validity of question wordings used to assess gambling expenditure. *International Journal of Social Research Methodology, 10*(1), 63–77.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Glenn W. Harrison[1,4,5] · **Morten I. Lau**[2,6] · **Don Ross**[3,4,5]

Glenn W. Harrison
gharrison@gsu.edu

Morten I. Lau
mla.eco@cbs.dk

1 Department of Risk Management and Insurance, Robinson College of Business, Georgia State University, Atlanta, USA

2 Copenhagen Business School, Frederiksberg, Denmark

3 School of Society, Politics, and Ethics, University College Cork, Cork, Ireland

4 School of Economics, University of Cape Town, Cape Town, South Africa

5 Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, Atlanta, USA

6 Durham University Business School, Durham University, Durham, UK