Disordered Gambling Prevalence: Methodological Innovations in a General Danish Population Survey

by

Glenn W. Harrison, Lasse J. Jessen, Morten Lau and Don Ross[†]

July 2017

Forthcoming, *Journal of Gambling Studies*

ABSTRACT.

We study Danish adult gambling behavior with an emphasis on discovering patterns relevant to public health forecasting and economic welfare assessment of policy. Methodological innovations include measurement of formative in addition to reflective constructs, estimation of prospective risk for developing gambling disorder rather than risk of being falsely negatively diagnosed, analysis with attention to sample weights and correction for sample selection bias, estimation of the impact of trigger questions on prevalence estimates and sample characteristics, and distinguishing between total and marginal effects of risk-indicating factors. The most significant novelty in our design is that nobody was excluded on the basis of their response to a 'trigger' or 'gateway' question about previous gambling history. Our sample consists of 8,405 adult Danes. We administered the Focal Adult Gambling Screen to all subjects and estimate prospective risk for Disordered Gambling. We find that 87.6% of the population is indicated for No Detectable Risk, 5.4% is indicated for Early Risk, 1.7% is indicated for Intermediate Risk, 2.6% is indicated for Advanced Risk, and 2.6% is indicated for Disordered Gambling. Correcting for sample weights and controlling for sample selection has a significant effect on prevalence rates. Although these estimates of the 'at risk' fraction of the population are significantly higher than conventionally reported, we infer a significant decrease in overall prevalence rates of detectable risk with these corrections, since gambling behavior is positively correlated with the decision to participate in gambling surveys. We also find that imposing a threshold gambling history leads to underestimation of the prevalence of gambling problems.

1. Introduction

We surveyed gambling behavior, and particularly disordered gambling behavior, in the general adult population of Denmark. The survey was administered in 2015 to provide precursor data, and to pre-identify subject strata for recruitment purposes, for a series of economic experiments with gamblers. The primary aim of our survey was therefore not to furnish comprehensive new prevalence data for the sake of informing Danish public health policy. Nor was our aim to be able to compare disordered gambling prevalence in Denmark with prevalence in other jurisdictions. The unusual nature of the motivations for our survey, in contrast with standard prevalence studies, gave us freedom to introduce a number of methodological innovations. Motivating and illustrating these is the immediate objective. Of course, the value of methods is ultimately determined by the usefulness and reliability of the results they yield. Thus we include full presentation of our results in the standard way. However, they are not our first-order focus of attention here.

As economists, our primary interest is in predicting and understanding the welfare consequences of policies and institutions. These consequences are not restricted to those that would be deemed relevant in a clinical context, though of course those are important. This wider focus of concern is the first general driver of our methodological innovations. A second general motivation is to generate data that facilitate experimental interventions with gamblers, both in the laboratory and the field. Analysis of experimental data depends on structural modeling of the risk preferences and beliefs of subjects that are inferred rather than directly observed. These inferences are based on gathering different information about our sample than is typically included in disordered gambling prevalence studies.

An explicit goal of much historical survey research on gambling disorder is to mimic prevalence estimates that would be obtained from face to face assessments by trained clinicians. This

objective also explains the revisions that have been made through the various editions of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*), which have in turn strongly influenced the design of gambling disorder screens. We discuss methodological limitations of these screens not because we doubt their value for estimating potential demand for clinical services, but because the screens are not, unsurprisingly, ideally adapted to all purposes, including those that motivate us. On the other hand, the relationship between results obtained using standard screens and other kinds of instruments is highly relevant, especially as clinically relevant harms from gambling are *among* the welfare consequences of economic concern even though they do not exhaust them.

In Section 2 we identify potential terminological confusions around 'problem,' 'pathological' and 'disordered' gambling that arise from the history of the literature, and particularly from the history of the *DSM*. In section 3 we review the methodological limitations of standard gambling disorder screens that our Danish survey was designed to transcend. Section 4 describes our survey design. In Section 5 we describe our results. Section 6 provides a concluding critical discussion.

2. Classification of Disordered Gamblers

Prevalence surveys of gambling problems have often aimed at binary classification of individuals as 'pathological' gamblers, or not, where these terms are defined either directly or approximately in terms of *DSM-IV* clinical criteria presented in American Psychiatric Association [1994]. Respondents reporting five or more of ten of these criteria have historically been so classified. *DSM-IV* listed Pathological Gambling among "Impulse Control Disorders Not Elsewhere Classified." In *DSM-5*, presented in American Psychiatric Association [2013] growing evidence of behavioral, etiological, and clinical response similarities between pathological gambling and addictive substance dependencies inspired its reclassification among "Substance-Related and Addictive Disorders," and to

its being re-named as "Gambling Disorder."

Introduction of this new terminology, more consistent with other *DSM* nomenclature, offers the prospect of correcting historical inconsistency in use of the term 'pathological gambling' as sometimes contrasted with, and sometimes used as synonymous with, 'problem gambling' (as noted by the Committee on the Social and Economic Impact of Pathological Gambling [1999]). It had been argued by some authors, including Petry et al. [2014; p.494], that 'pathological gambling' should be retired from professional use because it is associated with social stigmatization. However, the implication of this motivation, that 'gambling disorder' should replace 'pathological gambling,' has not become consensus practice.

Researchers analyzing data from recent waves of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), such as Pietrzak et al. [2007], Algeria et al. [2009] and Nower et al. [2013], following a sub-tradition of using 'problem gambling' to refer to the presence of 'pathological gambling' at pre-clinical levels, and as denoting an early or risk-indicating possible precursor stage to full-blown 'pathological gambling,' have applied 'pathological gambling' to the most severe manifestation of 'gambling disorder' and applied 'problem gambling' to less acute cases of 'gambling disorder.' These authors, to whom we will refer for brevity as 'NESARC analysts,' have applied a scoring rule to the *DSM*-based screen due to Fisher [2000], according to which a sub-clinical 'problem gambler' is someone who who has between one and four positive responses with at least one of these being to one of the final three *DSM-IV* criteria.[1] This policy leaves all three phrases in play, and uses 'gambling disorder' to denote a larger class of people, namely all respondents satisfying Fisher's sub-clinical threshold, than the *DSM-5* does.

---

[1] These are: (8) Have you been forced to go beyond what is strictly legal, in order to finance gambling or to pay gambling debts?; (9) Have you risked or lost a significant relationship, job, educational or career opportunity because of gambling?; (10) Have you sought help from others to provide money to relieve a desperate financial situation caused by gambling?

The usage interacts particularly confusingly with the measurement screen used most commonly in recent prevalence studies, the Problem Gambling Severity Index (PGSI) of Ferris and Wynne [2001]. In the PGSI, 'problem gamblers' are those comprising the highest severity category. Thus, in effect, though not in intention, the PGSI uses 'problem gambling' as an approximate synonym for the *DSM-5* category 'gambling disorder.' Thus, PGSI-designated 'problem gamblers' and 'problem gamblers' in the usage of the NESARC analysts are intended to be disjoint sets!

In light of this semantic chaos, researchers must stipulate terminological policies, particularly if, as we do, they use the PGSI. Henceforth, and aligned with usage in Harrison, Lau and Ross [2016], where we refer to the clinical phenomenon *ex cathedra* we will follow *DSM-5* and use 'gambling disorder' (GD). We will refer to a representative person who has acquired the condition as a 'disordered gambler' (DG). Where we refer to previous work set in clinical contexts that used either 'pathological' or 'problem' gambling without distinguishing them, or intending that they be distinguished (for example, in work applying the PGSI), we will anachronistically use the terms 'gambling disorder' and 'disordered gambler.' Given that *DSM* and PGSI threshold criteria are similar but not identical, this unavoidably creates potential ambiguity when studies using different instruments are compared, but it at least aligns intended references, so that the question of who is said to suffer from GD depends on empirical facts rather than semantic decisions. Where we refer 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, making this special departure from our policy explicit. As a factor that reduces scope for confusion, note that the threshold for being scored as a DG in *DSM-5* is the same as the threshold for being scored as a pathological gambler in *DSM-IV*. 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.

3. Methodological Issues in Research Applications of Gambling Disorder Screens

We consider several critical issues that arise with screens based exactly or approximately on *DSM-IV* criteria.

The first issue is that a DG is conceptualized as someone who would be diagnosed as such in a clinical interview if they presented themselves for treatment at a mental health clinic or other suitable facility. This is of course a legitimate indicator to want to detect, but is not the only basis for interest in the prevalence of gambling problems for public health forecasting, and for economic welfare analysis of the consequences of gambling policy. For these purposes we are also interested in identifying people with histories of GD who are not currently manifesting symptoms, but who might thus be vulnerable to re-occurrence under certain conditions, as well as people who may be regarded as at risk because they satisfy a number of *DSM* criteria but do not reach the diagnostic threshold. Such gamblers are sometimes referred to as 'sub-clinical' cases (e.g. Blanco et al. [2006]), but most screens other than the PGSI simply classify them with other respondents who are not DGs. This issue has direct implications for the understanding of 'risk for GD' as this is operationalized in survey screens and prevalence estimates. It is common for researchers to refer to people with positive sub-threshold scores as being 'at risk,' and the PGSI explicitly sorts sub-threshold respondents who gamble into the categories of "Low risk" and "Moderate risk."

The expression 'risk' here is not intended to refer, as it would in a standard public health prevalence estimation exercise, to the relative probability of developing GD in the future, either unconditionally or conditionally on measured factors. Rather, in the context of the clinical diagnostic benchmark, it refers to the probability that the respondent would currently be diagnosed as a DG in a longer, more probing clinical interview conducted by a medical professional. This special-purpose

understanding of 'risk' is unduly narrow, and problematic in having no forward-looking aspect, for purposes of public health forecasting and welfare analysis.

A second issue that all gambling survey instruments of which we are aware employ a 'screener,' 'gateway' or 'diagnostic stem' question to filter out individuals who have never engaged in gambling behavior beyond some pre-specified threshold. In many cases the filter is whether a respondent has ever lost a certain amount of money in any one day of gambling, and in other cases it is whether they have ever gambled more than a certain number of times in any year. Sometimes filtering criteria are combined. For example, the National Comorbidity Survey Replication (NCS-R) documented by Kessler et al. [2008] asks if the respondent has ever gambled 11 or more times in any one of a wide range of activities and ever lost $365 or more in a year; the NESARC has a gateway question that asks if the individual has ever gambled in any organized form 5 times in one year; and the Canadian Community Health Survey (CCHS) Cycle 1.2 asked, inter alia, "In the past 12 months, how often have you bet or spent more money than you wanted on gambling?" and excluded anyone that responded "I am not a gambler" (Cox et al. [2005]).

As Stone et al. [2015] have noted, this directly implies sample selection bias in analysis of survey results for purposes of prevalence estimation. Harrison, Lau and Ross [2016] compute the extent of this bias, using a correction method due to Heckman [1976][1979] that is a standard tool in econometrics, with the NESARC gambling data. Harrison, Lau and Ross [2016] find that the predicted mean of *DSM* scores in the sample increases significantly, as does the estimated prevalence of GD. One of the *DSM-IV* criteria for GD is that the respondent has lied to family or others to hide the extent of their gambling, but the respondent is assumed not to do this when asked if they gamble in a survey, particularly in response to the gateway question. The possibility that someone might lie in

response to the single gateway question in the *NESARC*,[2] to avoid being asked questions about their gambling behavior, is well documented in general clinical test-retest interview settings: see Kessler et al. [2004; p.125].

A final issue is that clinical criteria can interact with political incentives when applied in general population surveys (Sadler [2013]). It is not hard to discern from discussions of revisions to the *DSM* criteria for GD and clinically-based instruments that a recurring concern has been that prevalence estimates cannot be allowed to get 'too high,' because that would imply a shortage of funding for mental health treatment (Pierre [2013]). Clear (enough) statements of this 'tail wags dog' problem can be easily found, and the issue is well known in the research folklore.[3] In another important corner of the supply-demand nexus, operators in the gambling industry welcome prevalence measures for pathological gambling or gambling disorders that are very low. For instance, the *American Gambling Association*, an industry lobbying body, notes that "Although the vast majority of Americans are able to gamble responsibly, a small percentage of people – approximately 1 percent of the adult population – cannot."[4] These prevalence estimates come from measures of GD defined by clinical criteria. The thrust of the industry's standard rhetoric on responsible gambling is that 99 out of 100 gamblers are just having fun, and should be left alone, and regulators should only worry about the 1-in-100 who is not gambling responsibly.

---

[2] The question was, "Have you ever gambled at least 5 times in any one year of your life?"

[3] For example, Regier et al. [1998; p. 110] comment that "Both the scientific and political implications of these high prevalence rates were highlighted by the timing of this release during the national debate on health reform. Major policy questions were raised about the need for mental health services that were implied by these high rates, along with concerns about possible insurance cost-benefit consequences. Some major media commentators identified such high rates as indicating a bottomless pit of possible demand for mental health services." More recently, from Petry et al. [2014; p.497]: "The American Psychiatric Association requires strong empirical data in support of changes to *DSM-5* that would substantially increase the base rate of a disorder." But the only motivation then mentioned is the circular argument that reducing the number of threshold criteria would make the base rate increase.

[4] On http://www.americangaming.org/social-responsibility/responsible-gaming, accessed 9/21/2014.

Again, we stress that these concerns arise for a simple reason: the general purpose survey instruments were intended, by design, to mimic and correlate with the screening that would occur in a clinical setting. Whether or not that setting might be a 'gold standard' for some mental health screening purposes, it simply differs from other reasons for wanting to measure gambling problems. Society may take a broader view of what constitutes a gambling problem, and economists concerned with welfare estimation must certainly take a broader view.

Our primary objective, then, is to estimate the prevalence of gambling problems in a general population and evaluate surveys of gambling problems that did not restrict themselves to clinical criteria for gambling disorders. We want to see how well these measures correlate with traditional clinically-oriented surveys, of course, if for no other reason than to integrate use of these instruments into a wider constellation of scientific probes. At the same time, it should not be taken for granted that finding low correlations between prevalence of clinical GD and prevalence of other forms of welfare-reducing gambling behavior necessarily constitutes a problem.

Our secondary objective is to have a 'wide screen' for subjects to be recruited into subsequent experiments designed to evaluate behavior in controlled gambling tasks. Our view is that the litmus test for scientific value of these instruments is whether they can predict, or be reweighted or calibrated to predict, actual gambling problems, in the sense of behavioral patterns that predict welfare loss.

Our overall survey design is intended to contribute to answering several methodological questions. One, as noted, is the extent of correlation between different survey instruments of gambling behavior, particularly instruments designed to identify different latent factors relevant to gambling behavior. Another methodological issue is the value of randomization of question order, and in some cases instrument order. A third issue is whether one uses lifetime or only more recent (e.g., past year's) gambling behavior as the time frame for responses. A final issue is the role of the 'trigger'

or 'gateway' question.

4. Sample Frame and Survey Instruments for the Danish Study

We contracted with *Analyse Danmark* (http://www.analysedanmark.dk/english) to obtain

10,000 completed survey responses from the adult population of Denmark between 18 and 75 years of

age. This sample was to be assigned equally to all treatments. Our completed sample consisted of

8,405, which is 12.8% of the sample frame of 65,592 Danes contacted. Of those contacted that did

not complete the survey, 3,331 started but gave up before completing. The cost of these surveys was

272,425 DKK, which was just over US$45,400 at the time they were implemented.

The sample was stratified according to sex and age across three regions in Denmark: greater

Copenhagen, Jutland, and Funen & Zealand. We assigned different weights to the three regions, with a

50% weight on the sample from greater Copenhagen and a 25% weight on each sample from the two

other regions. This design allowed us to recruit subjects for later experiments from a relatively large

sample in greater Copenhagen. The respondents in our survey were recruited from two internet-based

panels with 165,000 active members.[5] Invitations were sent by e-mail and respondents could answer

the survey questions on the internet using personal computers, mobile phones or tablets. They were

told in the invitation letter that 40 respondents who completed the survey would be randomly chosen

to receive a gift card of 500 kroner.[6] Summary statistics for all participants and non-participants are

provided in Appendix C.

A primary instrument we adopted for use in Denmark that has not featured in past GD

prevalence studies is the Focal Adult Gambling Screen (FLAGS) designed by Schellinck et al.

---

[5] *Analyse Danmark* have a panel of 25,000 active members, and *Userneeds* have a panel of 140,000
members. The two internet panels are regularly updated and member are recruited via the internet (banners,
newsgroups, etc.), email, and by phone.
[6] The gift cards were issued by www.gavekortet.dk, an internet based portal for gift cards.

[2015a][2015b]. A first crucial design feature of the FLAGS relevant to our goals is that, unlike gambling screens derived directly or approximately from the *DSM* criteria, it seeks to identify risk of manifesting GD in the sense that is standard in public health surveys. That is, when the FLAGS sorts respondents into categories labeled "No Detectable Risk," "Early Risk," "Intermediate Risk," "Advanced Risk," and "Problem Gambler," risk is to be understood as referring to the probability that the respondent will develop GD at some time in the future (which might be immediate). This is in contrast to the interpretation of 'risk' intended explicitly by the designers of the PGSI, and often used in interpreting data from *DSM*-based screens, which denotes the probability that the survey score mis-predicts a current positive diagnosis that would be obtained through a clinical interview. A second, related, design feature of the FLAGS is that it is intended to measure both reflective and formative constructs associated with GD. A reflective construct is an indicator of a latent variable that is believed to partly constitute the reflected construct as manifest in behavior. A formative construct is one that is often, but not always, found along the etiological pathway to the target behavioral marker. Thus reflective constructs should show high inter-correlation, while no such restrictive expectation applies, independently of a specified structural model, to formative constructs.

Most of the major surveys of gambling behavior based on clinical criteria rely solely on reflective constructs. There have, however, been surveys of gambling propensity building on formative constructs, most notably Breen and Zuckerman [1999]. They administered their own Gambling Beliefs and Attitudes Survey (GABS) to college students, and then studied their actual gambling behavior in a laboratory card-game. A formal psychometric evaluation of GABS was undertaken by Strong et al. [2004a], boiling the original 35-item instrument down to a preferred 15-item instrument. The same general psychometric methods can in principle be used to re-tool clinically based instruments to identify a continuum of gambling types rather than just the binary classification into DGs and others:

see Strong et al. [2003][2004b] for an application to the South Oaks Gambling Screen (SOGS), Strong

and Kahler [2007] for an application to the Alcohol Use Disorder and Associated Disability Interview

Schedule – *DSM-IV* (AUDADIS-IV), and Sharp et al. [2012] for an application to the PGSI. The

efforts for SOGS and AUDADIS-IV did not meet with great success, suggesting that more

fundamental *ex ante* survey design methods are needed as a complement to *ex post* statistical forensics.[7]

The use of formative constructs in the FLAGS invites development of structural models, of the kind

favored by economists and econometricians, in analyzing findings based on it, as opposed merely to

application of simple statistical significance tests.

The latest version of FLAGS contains 64 questions designed to measure 10 latent constructs.

As applied to machine gambling, these constructs are:

1. **Risky Cognitions: Beliefs (RCB)**, such as irrational or inaccurate beliefs about machine gambling.
2. **Risky Cognitions: Motives (RCM)**, such as risky reasons for gambling (e.g., to pay off bills, to escape problems, for self-esteem or status).
3. **Preoccupation: Desire (POD)**, such as a strong drive to play the machines as much as possible.
4. **Impaired Control: Continue (ICC)**, such as the inability to stop playing slots/machines once started.
5. **Risky Practices: Earlier (RBE)**, such as less extreme types of risky practices that usually precede more harmful practices (e.g., using bank card to get more money to play).
6. **Risky Practices: Later (RBL)**, such as more extreme or harmful types of risky practices (e.g., using credit to finance play).
7. **Impaired Control: Begin (ICB)**, such as an inability to resist or stop oneself from going to play slots/machines.
8. **Preoccupation: Obsessed (POO)**, such as excessive preoccupation, constantly thinking about slot gambling or finding ways to gamble on machines.
9. **Negative Consequences (NGC)**, such as negative impacts in at least 3 of 14 different areas of life including financial, personal, family, work, health, social.
10. **Persistence (PST)**: such as continuing to gamble, over an extended period, in a risky manner that leads to harms.

Five of these constructs are regarded by the FLAGS designers[8] as formative (items 1, 2 5, 6 and 9),

---

[7] The inconclusive findings where the SOGS is concerned might also stem from the absence of a conceptual or historical relationship between the SOGS and *DSM* criteria for GD.

[8] Personal communication, August 24, 2016.

and the other five as reflective. The complete list of statements can be found in Appendix A.

We consider two other popular survey instruments for existing gambling problems. One is the 9-item PGSI. The second is based directly on the *DSM-IV* criteria for GD. Both the PGSI and the *DSM-IV* screen are intended to measure reflective constructs associated with GD.

Whether a construct should be regarded as formative or reflective is not independent of theory. Consider, for example, visceral cravings or urges to gamble. In one sense cravings are formative, in that their occurrence predicts GD but are not a necessary precursor to it. However, on the current standard general conceptualization of addiction (Ross et al. [2008], Redish et al. [2008]), the fact that a person experiences cravings is a very strong indicator, probably the most reliable indicator of all, that gambling has become addictive for that person (Redish [2009]). In this theoretical context, cravings are a reflective construct. Closely related to the cravings construct is that of preoccupation with gambling, the extent to which thoughts about gambling tend to crowd out efforts to devote attention to other activities. This ambiguity would necessarily be resolved in a structural model of propensity to manifest GD, as invited by the interpretation of risk that informed design of the FLAGS. The ambition to develop and empirically estimate such a model calls for gathering more information than is obtained from any existing survey screens, including the FLAGS. Our study incorporates two survey instruments aimed at cravings, the Gambling Craving Scale (GACS) developed by Young and Wohl [2009], and the Gambling Urge Screen (GUS) developed by Raylu and Oei [2004a]. We additionally use one instrument aimed at detecting preoccupation, the Gambling Related Cognitions Scale (GRCS) developed by Raylu and Oei [2004b]. Appendix B explains how each instrument was scored. In most cases the scores follow the standard algorithms, but in some instances the scoring is not obvious.

In addition, we asked several questions to identify past gambling behavior, but we did not use

these as the basis for an exclusionary 'trigger' question. In separate questions, each subject was asked if they had lost more than 40 kroner or 500 kroner on gambling in a single day. The lower amount corresponds to the price of a common Danish state lottery ticket, and the larger amount to a naturally larger denomination that a typical Dane would likely recall. We asked these questions in the lifetime frame for 50% of the sample and in the time frame that spanned the previous 12 months for the other 50% of the sample. These questions allowed us to directly analyze the consequences that would have obtained for prevalence estimates if they had been used as exclusionary triggers.

There is a long history of interest in mental disorders that sometimes or often co-occur with GD. Indeed, some research points to virtually every measured psychiatric disorder as being correlated with GD.[9] The implication of much of this research is to focus attention on common causes of several psychiatric disorders. Our evaluation cannot be that exhaustive, but we do consider instruments to measure a number of conditions that have been most frequently reported as comorbidities. These survey instrument are the Beck Anxiety Index (BAI) developed by Beck et al. [1988] and Beck and Steer [1990], the Beck Depression Inventory (BDI) developed by Beck et al. [1961], and the Barratt Impulsivity Scale (BIS) developed by Patton et al. [1995].[10] Because alcohol abuse and/or dependence is often reported as highly correlated with GD, and of particular concern in Denmark, we implemented the Alcohol Use Disorders Identification Test (AUDIT) of Babor et al. [2001]. We also asked respondents whether they currently smoked, and if so how many cigarettes per day.

To allow for the effects of recent life events, we asked if the individual had experienced the

---

[9] For instance, evaluating data from the NESARC Petry et al. [2005; Table 3, p. 570; model 3] report lower bounds of 95% confidence intervals of Odds Ratios in excess of 1 for alcohol dependence, any drug abuse, any drug dependence, nicotine dependence, major depressive episodes, dysthymic disorders, manic episodes, panic disorders, social phobia, specific phobia, generalized anxiety, and every personality disorder considered (avoidant, dependent, obsessive-compulsive, paranoid, schizoid, histrionic and antisocial). Kessler et al. [2008; Table 2, p.1357] report a similar list from the NCS-R.

[10] Major Depression and Anxiety are regarded as Axis-1 mental disorders. Impulsivity, in the psychiatric literature, is constructed as a personality trait.

death of an immediate family member (partner, child, parent or sibling) in the past 12 months, or had been hospitalized for a major medical problem during the past 12 months.

To evaluate the effects of order we only evaluated one instrument along with FLAGS for any one person, and randomized the order of presentation. We split our sample into 10 treatments: (1) **FLAGS**, PGSI, BIS; (2) PGSI, **FLAGS**, BIS; (3) **FLAGS**, *DSM-IV*, BAI, AUDIT; (4) *DSM-IV*, **FLAGS**, BAI, AUDIT; (5) **FLAGS**, GACS, AUDIT; (6) GACS, **FLAGS**, AUDIT; (7) **FLAGS**, GUS, BDI, AUDIT; (8) GUS, **FLAGS**, BDI, AUDIT; (9) **FLAGS**, GRCS, AUDIT; and (10) GRCS, **FLAGS**, AUDIT. The only difference between the odd and even treatments here is the order of the gambling instrument, to assess if comparisons of FLAGS with the other instruments are affected by participants having already completed the more expansive FLAGS.[11]

We further split each treatment equally into cases in which the timeframe for the gambling instruments FLAGS, PGSI and *DSM*-based screen are lifetime or past 12 months.

For 50% of participants we randomize within each thematic block, when possible, and otherwise present the questions in the standard order.[12] The software used to implement the survey did not allow randomization within a block unless the response formats were all the same, so we could not randomize the order for AUDIT, the BDI, and our few concluding questions.

5. Results

We are interested in answering several questions through analyses of these data. First, what is the distribution over the Danish adult population of GD indicators as assessed using the FLAGS instrument? The answer to this question provides the sampling frame for subsequent experimental

---

[11] There is a natural aggregation of these 10 sub-blocks into three groupings, often used in the field implementation of FLAGS (§1: RCB, RCM and POD, §2: ICC, RBE and RBL, and §3: ICB, POO, NGC and PST). We do not randomize across those groupings but do randomize within each grouping.

[12] In the development of FLAGS by Schellink et al. [2015a; p.149] the original 132 statements were randomized.

evaluation of actual gambling behavior by individuals recruited into our incentivized experiments. We want to evaluate the raw distribution, based on the sample that completed our surveys, since that is the basis that is typically used to assess population GD risk (in what we have called the 'standard public health sense' of risk).

Second, what is the inferred population distribution after correcting for sample weights and sample selection? The correction for sample weights, based on observable differences in the demographic mix of the sample and the population, is familiar in many survey settings, but is not always applied in assessments of gambling problems.[13] The correction for sample selection, based on unobservable differences of the sampled and non-sampled population, has never been applied in published assessments of GD prevalence in any population.

Third, how is the distribution of GD indicators affected by the treatments we considered? Does it matter if FLAGS comes first or second among gambling screens, or if we randomize question order? And how does the lifetime frame affect the distribution of GD indicators?

Fourth, what is the effect on inferences about the distribution of GD indicators of applying a threshold trigger question based on past gambling history? These trigger questions are usually applied *ex ante* the administration of the survey, and some gambling status then assumed for the individual. Our design deliberately avoided such assumptions, allowing us to impose them *ex post* the administration of the survey to study their effect.

Finally, what are the correlates of GD indicators, as assessed by FLAGS? We examine the correlation between different instruments, and when possible the partial correlation holding constant the effect of a third instrument.

---

[13] Sample weights are often *used* by economists when they are provided with data from well-designed surveys, but are rarely *constructed* by economists or gambling researchers in our experience. Harrison, Lau, Rutström and Sullivan [2005; §4.1] provide an exposition of procedures that can be used to construct sample weights for simple surveys. More complex sample designs require more sophisticated methods to construct sample weights: see Levy and Lemeshow [2008; ch. 16].

A. GD Indicators in the Danish Sample

Table 1 shows the distribution of GD indicators in the Danish sample based on the FLAGS instrument, applying the standard past-12-months frame, without application of any gambling history threshold. We find that 79.7% of the sample is, in terms of FLAGS response categories, at No Detectable Risk for GD, 12% is at Early Risk, 3.9% is at Intermediate Risk, 3.3% is at Advanced Risk, and 1.1% is classified as Problem Gamblers. Out of 8,405 in the sample, we detect 95 respondents who would, on the intended interpretation of FLAGS data, be regarded as probable current DGs.

Table 1 also shows the distribution of FLAGS outcome categories broken down by comparison with the 'risk' levels explicitly identified by the PGSI and sometimes used in interpreting *DSM* screen results (e.g. Stone et al. [2015]). The time frame in each case is the past 12 months. It is important to be reminded that this is not a comparison of matched constructs, except with respect to those identified as probable current DGs. 'Risk' in the PGSI assessment means 'risk of being a current false negative,' whereas 'risk' in the FLAGS means 'risk of manifesting GD at some future time (including the immediate future).' No concept of 'risk' is operative when the standard, binary *DSM*-based screen assessment is applied.[14]

The PGSI and *DSM*-based screen samples shown in Table 1 are smaller, at 1,671 and 1,757 respectively, than the FLAGS sample, but the former were assigned at random within the complete sample. For the *DSM* comparison, the biggest difference in classification in percentage terms is for those that FLAGS classifies as current DGs. Notably, this is precisely the category on which the

---

[14] As discussed earlier, when more refined categorizations are applied to *DSM*-based screen assessments, as in Gerstein et al. [1999], Fisher [2000], and Stone et al. [2015], the categories other than the designation as a current probable DG are sometimes interpreted by reference to 'sub-clinical risk.' This suggests that what is being measured is relative severity of symptoms. However, as with PGSI assessment, 'risk' refers not to risk of subsequently developing a more severe level of GD, but rather to risk that observation of lower symptom severity will lead to current false negative diagnosis.

screens are 'intended' by their respective design principles to agree with one another. Although the

sample of 12 is small, the *DSM*-based screen classifies only 1 of the 12 as a Pathological Gambler. The

other *DSM*-screen mismatch is for those that FLAGS classifies as being at Advanced Risk: *DSM*

classifies 98% (55 individuals) of those 56 individuals as being Non-Gamblers. The PGSI has a better

match with FLAGS on DGs (i.e., "Problem Gamblers" in the PGSI). However, of the 47 that FLAGS

classifies as being at Advanced (prospective) Risk, 27.7% are classified by PGSI as being Non-

Gamblers. Similarly, for those 62 individuals that FLAGS classifies as being at Intermediate

(prospective) Risk, the PGSI classifies 43.6% (27 individuals) as being Non-Gamblers. The number of

subjects identified by the FLAGS as being at some prospective risk for GD is significantly larger than

the number of subjects identified by the PGSI as being at risk of false current negative diagnosis. This

highlights one of the potential problems we identified with using a clinical screen as a research

instrument if one's objectives are derived from interest in public health forecasting or economic

welfare assessment.

B. *Disordered Gambling Propensity in Denmark*

We focus next on GD propensity in Denmark, as indicated by our data, and address two

questions. Can we identify individual characteristics that are correlated with prospective risk for GD as

operationalized by the FLAGS? What is the inferred distribution of this risk in the population after

applying sample weights and controlling for endogenous sample selection into the survey? We

estimate a Semi-Nonparametric Ordered Response model with and without sample weights and

corrections for sample selection.

Sample weights are constructed from administrative data at Statistics Denmark on the

population size of men and women in various age groups and regions in Denmark, and correct sample

estimates for overrepresentation of respondents by age, sex and region in the sample relative to the population.

We correct for **sample selection bias** using full information maximum likelihood estimation of the Ordered Response model, and follow the direct likelihood approach originally due to Heckman [1976][1979]. One important assumption in the 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 the survey question elicited a binary response, and both equations are modeled with probit specifications, for example, the natural first assumption is that the errors are bivariate normal. 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.[15] We model the second (main) equation as an ordered response, instead of a binary response, to reflect the FLAGS categories we are interested in.[16]

The Ordered Response models control for treatment variables and demographic characteristics such as sex, age, income and smoking behavior. Indicators for trigger questions (applied versus not applied) and subcontractor (*Analyse Danmark* or *Userneeds*) are also added as control variables in the models. Appendix E documents the detailed estimates for each model. Table 2 shows

---

[15] 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 (ordered) probit specification.

[16] It is common to refer to different levels of severity of psychiatric disorders. For instance, the American Psychiatric Association [2015; p.586] discusses GD as follows, referring to *DSM-5* criteria: "Severity is based on the number of criteria endorsed. Individuals with mild gambling disorder may exhibit only 4-5 of the criteria, with the most frequently endorsed criteria usually related to preoccupation with gambling and 'chasing' losses. Individuals with moderately severe gambling disorder exhibit more of the criteria (i.e., 6-7). Individuals with the most severe form will exhibit all or most of the nine criteria (i.e., 8-9)."

predicted risk levels from various models, to show the effects of sample weights and sample selection corrections.

In model A we estimate gambling risk without using sample weights and without corrections for sample selection. The predicted distribution of prospective risk for GD in the adult Danish population in the panel of Table 2 without using sample weights and controls for sample selection, but controlling for demographic characteristics and treatments, corresponds to the distribution reported in Table 1: 79.7% are indicated for No Detectable Risk, 12% are indicated for Early Risk, 3.9% are indicated for Intermediate Risk, 3.3% are indicated for Advanced Risk, and 1.1% are indicated as Disordered Gamblers. In effect, the statistical model is regenerating the raw sample data faithfully.

Model B in Table 2 employs sample weights, to be applied to the sample of 8,405 that completed our survey. Allowing for sample weights has very little impact on the prevalence of GD risk levels in the population.

Model C in Table 2 controls for endogenous sample selection as well as sample weights. We find a significant association between response rates and age groups: compared to those aged between 30 and 39, younger subjects have lower response rates ($\approx$ 2.3 pp on average), ripe aged subjects between 40 and 49 have higher response rates ($\approx$ 2.0 pp on average), and subjects older than 50 have much higher response rates ($\approx$ 13.1 pp on average). Women are significantly more likely to respond ($\approx$ 1.1 pp on average). Both of these demographics, age and gender, point to our sample selecting on the basis of characteristics that are known *a priori* to correlate with propensity to gamble and be at risk for GD, but in opposite directions.

We also find a significant *positive* correlation of 0.21 between the error terms in the main equation and in the selection equation. This result indicates that people with higher gambling risk are also more likely to participate in the survey, and *this* selection effect is due to *unobservable characteristics*

that are not captured by the demographic variables in the selection equation and the main equation.

The implication of these findings from the sample selection estimation is that we infer a much *higher* prevalence of No Detectable Risk (87.6% compared to 79.2% in model B). Hence we infer a *lower* prevalence of 'at risk' Danes once we correct for sample selection.

A particularly interesting result of correcting for sample selection is that the *mix* of 'at risk' Danes changes. With correction for sample selection we infer a sizeable reduction in the probability of being indicated for Early Risk (5.4% versus 12.1% in model B), a reduction of just over 2 percentage points in the probability of being indicated for Intermediate Risk (1.7% versus 3.6%), a reduction of just over 1 percentage points in the probability of being indicated for Advanced Risk (2.6% versus 3.8%), and an increase in the probability of being indicated for Disordered Gambling (2.6% versus 1.3% in model B).

Figures 1 through 5 display the marginal effects of important covariates on the probability of each GD risk level, taking into account the sample selection effects as well as the probability of specific risk levels. The dots in Figure 1, for example, show the point estimates of the effect of being female, and the shaded bars show the 95% confidence interval, making it possible to see if the effect is statistically significant. The vertical axis is the change in probability, so a value of +0.02 indicates a change of 2 percentage points. Figure 1 shows that being a female significantly increases the probability of being classified as having no detectable risk, and correspondingly decreases the probability of being classified at different risk levels. Although the prediction for any one of the 'at risk'' levels is not statistically significant, since the upper bound of the 95% confidence intervals just include zero, the joint effect is statistically significant.[17] Figure 2 shows that lower income individuals tend to be more at risk of GD, although the effects are not statistically significant at the exact 95%

---

[17] This is implied by the fact that the 'at risk' levels are jointly complementary to the 'no detectable risk' level, which is statistically significantly different from zero.

level.[18]

C. Comparisons

We can compare our results to previous Danish gambling prevalence studies by Bonke and

Borregaard [2006][2009] and Ekholm et al. [2012]. The study by Bonke and Borregaard [2006] includes

sample weights but does not control for sample selection bias, while the other two studies do not

control for sample weights or sample selection bias. We therefore compare our uncorrected GD

prevalence rates with the prevalence rates they report.

Bonke and Borregaard [2006][2009] used the National Opinion Research Center *DSM*

(NODS) screen to estimate prevalence of GD in a sample of 8,153 Adult Danes between 18 and 74

years of age.[19] The NODS, like the interpretation of *DSM*-based screens developed by Fisher [2000],

uses a pre-clinical 'Problem Gambling' construct that has no counterpart in the FLAGS and the PGSI.

Like other screens that treat accurate clinical diagnosis as an epistemological gold standard, it also uses

classifications of 'risk' that refer to probabilities of inaccurate prediction of current clinical diagnosis

rather than propensity to develop GD. In light of these issues, it is only meaningful to compare our

(uncorrected) GD prevalence prediction with the Pathological Gambling prevalence prediction of

Bonke and Borregaard [2006][2009], and only for the past-year frame.[20] Their sample frame of 11,737

people was randomly drawn from the Danish Central National Register and stratified according to sex,

---

[18] We do not dismiss results that cross the trip-wire of a 95% confidence interval, and simply avoid referring to them as statistically significant.

[19] The NODS was developed by Gerstein et al. [1999] and builds on the *DSM-IV* gambling screen. It probes 17 reflective constructs with questions that measure lifetime and past-year risky gambling behaviors. Bonke and Borregaard [2006] compare the NODS instrument with the SOGS in a pre-test with 1,232 subjects and find that the NODS detects a lower prevalence of GD.

[20] Bonke and Borregaard [2006][2009] also estimated prevalence based on a lifetime frame. Given the interpretation of 'risk' at work in the NODS, the lifetime frame prevalence risk estimate should be understood as referring to the probability that a respondent would have been falsely negatively diagnosed at some time in his or her life. In light of the FLAGS risk concept being prospective, there is no corresponding interpretation of any FLAGS measure.

age, geographical information and marital status. Their survey was conducted mainly by telephone and in some cases by face-to-face interviews, and the overall response rate was 69.5%. They identify 0.26% of their sample as being Pathological Gamblers. This is significantly lower that our (uncorrected) FLAGS-based estimate of 1.1%. All of the respondents of Bonke and Borregaard [2006][2009] were first asked a trigger question on whether they had ever lost 35 kroner in a single day. They were classified as non-gamblers without administering the full survey if they answered "No." Bonke and Borregaard [2006][2009] report higher prevalence rates of Pathological Gambling for men than for women, lower prevalence rates for respondents older than 45, and lower prevalence rates for respondents in the highest income quartile than in lower quartiles. They find no significant effect of using sample weights calculated for sex, age, region, and marital status on estimated prevalence rates. Despite having access to administrative data for the full sample frame, the estimated coefficients by Bonke and Borregaard [2006][2009] are not corrected for endogenous sample selection, hence the actual prevalence rate for the population may be different from their estimate for the sample.

Ekholm et al. [2012] used data from the Danish Health Interview Survey in 2005 and the Danish Health and Morbidity Survey in 2010. The two samples were randomly drawn from the Danish Central National Register and included 10,916 respondents in 2005 and 23,405 respondents in 2010. After the main face-to-face interview the respondents were asked to complete a self-administered questionnaire that, among other things, included two questions that are related to gambling behavior. The so-called lie/bet questionnaire, which consists of two questions from the *DSM-IV* screen, was used in the survey, and those respondents who answered yes to at least one of the two questions are classified as DGs.[21] The final sample contains 5,686 respondents in 2005 and 14,670 respondents in 2010. Ekholm et al. [2012] find that the prevalence rate of past-year GD is

---

[21] The two questions are: "Have you ever lied to people important to you about how much you gambled?" and "Have you ever felt the need to bet more and more money?" The possible answers were: Yes, in the past 12 months; Yes, previously; No, I never gamble.

0.9% in 2005, and falls to 0.8% in 2010.[22] This is significantly lower than our FLAGS-based estimate

of 1.1%. Prevalence rates are higher among men than women and decrease with age. They do not

report any estimates that use sample weights or correct for endogenous sample selection, but mention

that they find no evidence of selection bias in unreported estimates.[23]

D. Effects of Treatments

The effects of treatments on gambling risk levels determined by FLAGS can be gauged by

standard measures of association applied to a 5×2 contingency table, but for more informative analysis

we use the SNP Ordered Response statistical model. The statistical model allows us to estimate the

size and significance of the effect of a variable on the probability of each level of prospective risk for

GD. It also allows us to examine the marginal effect of the treatment, controlling for other correlated

effects. For some inferential purposes we want to know the total (unconditional) effect, but typically

we are interested in the marginal effect.

We find that when FLAGS is presented first it does increase the likelihood of someone being

classified as having a detectable risk, particularly as being indicated for Early Risk.[24] The opposite

qualitative effect occurs when we randomize question order within the FLAGS instrument.

Randomization significantly lowers the likelihood of being classified with a detectable risk, again with

the biggest effect on indication for Early Risk.

We analyzed the use of a lifetime frame for FLAGS questions as a treatment. This can be

understood as interrogating the interpretation of 'risk' as prospective that was intended in the design

---

[22] Ekholm et al. [2012] also estimated lifetime-frame prevalence, which has the same interpretation as in Bonke and Borregard [2006][2009], and again has no counterpart measure in a FLAGS lifetime frame.

[23] Ekholm et al. [2012; p. 8] discuss whether differences in response rates between 2005 and 2010 may affect the results. They mention that "non-response adjusted prevalence estimates did not indicate that non-response bias affects the conclusion of the present study (data not shown)."

[24] A Pearson $\mathcal{X}^2$ test of the hypothesis of no association has a *p*-value less than 0.001.

of the FLAGS; it amounts to asking what information about current prospective risk of developing

GD is carried by a person's ever having manifested FLAGS indicators.[25] One might have expected *a*

priori to find indication of greater prospective risk for GD, since no one can have revealed fewer

indications of gambling problems over their lifetime than they have over the past year alone. However,

we find a significant and large reduction in detectable risk using the statistical model.[26] We implement

the lifetime gambling frame by asking the respondents to "…think about your lifetime gambling

experiences," as opposed to "…think about your gambling experiences in the last year." It is likely that

some respondents answered "yes" to a question when they were asked to consider their behavior in

the last year and "no" when they were asked to consider their lifetime behavior. For example,

respondents were asked to consider the following question "I would like to gamble almost every day"

and answer "yes" or "no." It is possible that respondents interpreted the lifetime frame ambiguously,

as excluding the very recent past (and in particular the past year). Thus someone who recently has

developed a gambling problem might answer "yes" when the question is framed over the past year and

answer "no" when the question is framed over the lifetime. It is also possible that a serious, episodic

gambling experience is recalled when a person is asked to reflect on the previous year, but not when

asked to reflect over a lifetime.

E. *Effects of Trigger Questions Based on Gambling History*

Our design allows an immediate data-only comparison of the effects of using trigger questions

based on gambling history to make inferences about prospective risk for GD. The usual tabulations do

---

[25] This value of setting FLAGS questions in a lifetime frame is thus conceptually distinct from the value of using lifetime frames for screens that understand 'risk' as probability of incorrect false diagnosis.

[26] A Pearson $\mathcal{X}^2$ test of the hypothesis of no association has a *p*-value less than 0.001. This is a two-sided test, but the direction of the effect is in the opposite of the alternative hypothesis that lifetime risks can be no smaller than recent risks.

not do justice to the careful language used in scoring FLAGS when gambling history is used. An individual is classified as a Non-Gambler in FLAGS if the threshold gambling history is applied, with this explanation of that category by Schellinck et al. [2011]:

> FLAGS instrument categorizes a person's risk based on their perceptions about and behaviors associated with gambling. It cannot therefore categorize a person's risk if they do not have gambling experience within the last year. There is a long list of correlates that have been shown to be associated with risk of problem gambling that we have left out of FLAGS that if possessed by an individual could indicate risk for problem gambling should they start to gamble. It was decided that in order to keep the instrument to a reasonable size its constructs would only be gambling specific; from the point of view of FLAGS these risk factors are therefore latent or unobservable.

This is saying that one *could* develop a different instrument if the intention was to consider gambling history as a determining factor for prospective gambling risk. On the other hand, FLAGS contains many more formative constructs suited to detecting latent risk than the clinically-referenced screens. Thus the above statement is exactly correct, and well stated in terms of latent and unobservable tendencies. It can be understood as pointing to the potential value of complementing the FLAGS with probes aimed at gathering data to inform development of a structural model of latent vulnerability to GD.

The same qualifications apply to the FLAGS construction of No Detectable Risk. This is defined as follows by Schellinck et al. [2011]:

> Those at No Detectable Risk do not flag on any of the risk indicators although it is possible that they answered yes to one or more statements making up some of the constructs. For those who answered yes to at least one statement there was insufficient certainty for us to say there was an indication on one of the dimensions. These people may still have unobservable or latent characteristics that would make them susceptible to becoming a problem gambler should the right conditions exist.

Again, the emphasis is on latent tendencies to develop and exhibit GD, which we seek to measure.

Thus when we study the effects of the trigger question on our FLAGS-based estimates, and find these effects to be significant, this should not be interpreted as criticism of the FLAGS design.

Instead, it should be understood as motivating development of complementary techniques for extending insight into latent risk indicators for GD, which could and should influence public forecasting and economic welfare assessment of gambling policies.

We find dramatic effects of imposing a gambling history threshold on administration of FLAGS questions. Figure 6 displays the results.[27] With the 500 kroner and 40 kroner threshold applied, we overstate the fraction of Danes that have no detectable risk, and understate those that have detectable prospective risk for GD. The bottom panel of Figure 6 rescales just on the detectable gambling risks, to provide more information on these effects. We find that 95% of the sample of 8,405 had not lost 500 kroner in gambling in one day, and that 18.1% of those had *some* detectable risk. Certainly those that had lost 500 kroner from gambling had a higher likelihood of exhibiting some prospective risk for GD according to FLAGS, as one would expect. But the crucial point is that it is *not* the case that individuals with *no historical gambling losses* of 500 kroner can be safely assumed to have *no detectible risk*, nor can those individuals *with* these losses be assumed to have *some* detectable risk. We stress again that the FLAGS classifications, properly interpreted, are clear on this point. Out of the complete sample, 46 individuals are indicated by the FLAGS instrument to be DGs, but did not say that they had lost 500 kroner in the past year. This is almost exactly the same number of individuals (49) indicated as DGs who *did* say that they had lost 500 kroner.

The 40 kroner threshold has a predictably smaller effect, since many Danes would have met this threshold compared to the 500 kroner threshold. In fact, 3,206 of the sample say that they had lost 40 kroner, compared to only 414 saying that they had lost 500 kroner. The fraction (38%) is less than a majority in the sample.

Figure 7 shows the effect of the thresholds for the classification of risk levels using the PGSI.

---

[27] Table F1 in Appendix F tabulates the detailed results.

Given the intended interpretation of 'risk' in the PGSI, which is focused on estimating the extent of currently undiagnosed GD in a population, significant effects are of greater critical concern. [28]

F. Correlates of Disordered Gambling Risk

Figure 8 shows the unconditional correlations of the FLAGS scores with the scores on other GD screens (*DSM*-based and PGSI), other gambling-related instruments (GACS, GRCS and GUS), an instrument measuring alcohol abuse and dependence (AUDIT), and instruments measuring anxiety, depression and impulsiveness (BAI, BDI and BIS, respectively). All correlations are statistically significantly different than zero at *p*-values less than 0.001, except the correlation with the BIS, which has a *p*-value of 0.094.

Where the GD screens are concerned, FLAGS is more strongly correlated with PGSI than with the *DSM*-based instrument, although both have positive and large correlation coefficients. The correlation of FLAGS with indicators of cravings, GACS and GUS, respectively, is very high, but the correlation with GRCS, which focuses on cognitive confusions about gambling and probability, is quite low. The correlation coefficient between FLAGS and the substance abuse instrument for alcohol is very low, as are the pairwise correlation coefficients with the measures of anxiety, depression and impulsiveness.

We find a different pattern of correlation when we examine each of the instruments in the context of our statistical model of the determinants of FLAGS prospective risk levels. This model controls for all the observable demographics and treatments considered earlier: we simply add the score or level of the instrument being studied and re-estimate. In each case the detailed impact varies with the FLAGS risk level, but the pattern is by now familiar from other marginal effects considered.

---

[28] As mentioned previously, the the *DSM*-based instrument we used was designed for a binary classification with respect to GD, so no interpretation of 'risk' is appropriate for it. Hence we do not display a counterpart to Figure 7 for the *DSM*-based screen assessment.

The impact on the No Detectable Risk level is the opposite sign as the impact on the 4 detectable risk levels, and the largest impact on a detectable risk level is for Early Risk.

Results for these marginal effects from the statistical model are shown in Appendix D. In summary, they show the predicted effects in terms of direction: someone that scores more highly on the *DSM*-based screen or PGSI also scores more highly on FLAGS. They also show much higher connections between the formative constructs of anxiety and depression than the unconditional correlations discussed above, and there is now a positive association between prospective risk for GD and the measure of impulsivity.

6. Discussion

A. Bias from Gateway Questions

A natural question is whether there is any comparison between gateway questions that would suggest that one is more or less biased than the other? We measure bias in terms of the effect on inferred prevalence of all 'at risk' categories. Harrison, Lau and Ross [2016] discuss several large-scale gambling surveys that employed extremely 'light' gateway questions that only exclude from the sample those that have *never* engaged in any gambling over some period, including the mere purchase of a lottery ticket.

One example is the *British Gambling Prevalence Survey* (BGPS) of 2010, which asked PGSI and *DSM-IV* questions for 73% of their entire sample of 7,756: see Wardle et al. [2011]. Harrison, Lau and Ross [2016; Figure 6] show that although there are statistically significant differences in prevalence estimates when correcting for sample selection bias using the *DSM-IV*-based screen, they are not quantitatively significant for policy purposes. Another example is the first wave of the *Victorian Gambling Survey* (VGS) of 2008, which asked PGSI questions for 75% of their entire sample of 15,000: see Billi et al. [2014][2015] and Stone et al. [2015]. Again, Harrison, Lau and Ross [2016; Figure 7] show that although there are statistically significant differences in estimated overall prevalence when correcting for sample selection bias using the PGSI, the differences for the highest risk categories (in the PGSI sense of 'risk') are not 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 analysis of VGS data illustrates the problems attendant on using *additional* threshold questions in order to reduce respondent time during the interview.[29] Starting with the 75% of the sample that had gambled at all in the last 12 months, these surveyors added 1,057 individuals who had gambled before then, to arrive at a lifetime sample of gamblers of 81%. But Stone et al. [2015] then employed an additional pre-screening procedure when applying the NODS screen to measure lifetime GD prevalence. This procedure asks 5 questions about gambling behavior, and only follows up with the additional questions of the full NODS instrument if someone responds

---

[29] 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 for each of the 13 activities. And then 24.0% of the sample was not asked the full set of CPGI questions because they said that they were a non-gambler on the first CPGI question. There were 98 subjects that refused to answer the initial questions about gambling activity, resulting in only 1,759 being asked the full set of questions and having any chance of being scored as "at risk." These deviations from the CPGI screen, and the PGSI index derived from it, were "approved by the authors of the scale" (Statistics Canada [2004; p.19]).

affirmatively to one of those 5 pre-screening questions. This procedure drops the VGS sample by 11,075, so that one ends up with only 8.5% of the full sample being evaluated with the NODS instrument. Unfortunately, this procedure is not statistically innocent: as Harrison, Lau and Ross [2016; Figure 8] show, it leads to a statistically and quantitatively significant sample selection bias when inferring lifetime prevalence. The fraction of the Victorian adult population that is indicated as being at risk of gambling disorders 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%.

So all that these results show is that there is no obvious way to ask gateway questions, consistent with the motivations for using them, that reduce bias across survey contexts. An alternative approach is to explore ways of asking different gateway questions to different subjects in order to better identify the effects of sample selection bias. For instance, one might include a sub-sample with 'light' gateway questions, or a sub-sample offered a monetary inducement to answer all questions without a gateway question.

B. Randomization Bias

One possible limitation of our approach is that we may have over-recruited gamblers because the incentive for recruitment was effectively a gamble: the chance of winning up to 500 kroner. This concern raises a more general issue known as 'randomization bias.' If subjects for surveys or experiments are recruited with a mix of a non-stochastic and stochastic fees, then there can be a sample selection bias with respect to subjects that are more or less risk averse. Higher non-stochastic fees encourage more risk averse subjects, *ceteris paribus*, to participate. On the other hand, higher stochastic fees encourage less risk averse subjects, *ceteris paribus*, to participate. In each case 'more' and 'less' is relative to the underlying population that one wants to make inferences about.

Apart from the immediate significance for estimates of population prevalence of GD, this issue is of fundamental methodological importance when using randomized treatments that subjects know about, or suspect are being applied. It is well-recognized in the older literature, [30] and has been demonstrated to be empirically relevant with controlled experiments by Harrison, Lau and Rutström [2010] and Harrison, Lau and Yoo [2014]. With appropriate variations in non-stochastic and stochastic recruitment fees, and a measure of risk attitudes for the sample, one can constructively measure and control for this bias.

C. Latent Vulnerability to GD

Detailed analyses of the statistical interactions of the elements of these related instruments, in this and other samples, furnish an empirical basis for the major project and methodological innovation of building a structural model of latent vulnerability to GD.

This objective merits some explanation, especially for non-economists. Gathering information about gambling behavior by directly asking subjects about it involves reliance on a restricted source that is likely to be systematically biased. In effect, it involves implicitly inferring patterns in actual gambling behavior from subjects' own conscious awareness of this behavior, which is likely to be less than complete, and inaccurate in some respects. We should also expect self-reports to often be self-exculpatory, or to reflect aversion to cognitive dissonance concerning a respondent's self-image.

In addition, researchers are not always in a position to be able to recruit large groups of gamblers, which include adequately-powered subsamples of potential DGs, to complete long surveys

---

[30] There is some evidence from the field of clinical drug trials that persuading patients to participate in randomized studies is much harder than persuading them to participate in non-randomized studies (e..g., Kramer and Shapiro [1984; p.2742ff.]). The same problem applies to social experiments, as evidenced by the difficulties that can be encountered when recruiting decentralized bureaucracies to administer random treatments (e.g., Hotz [1992]). For example, Heckman and Robb [1985] note that the refusal rate in one randomized job training program was over 90%.

such as the FLAGS. One might instead only be able to administer a shorter screen such as the PGSI, with all of the methodological limitations we have identified. Such epistemic limitations can be partly mitigated if one has an evidence-based model of general statistical relationships between FLAGS responses and PGSI responses that go beyond mere correlations. Data on which patterns of response to both screens are conditioned can provide crucial mediating leverage for this.

Self-report surveys that are only about gambling are poor instruments for revealing information about underlying causal factors that aren't themselves gambling behaviors. Though there are multiple paths to GD and to gambling problems more generally, and likely more than the three clusters identified in the 'pathways' model of Blaszczynski and Nower [2002], to suppose that each DG is a purely idiosyncratic case would be tantamount to believing that no general account of the causes of gambling problems is possible at all. Again, however, we can make progress toward such a theory only by systematically analyzing gambling survey data in conjunction with other data about gamblers and their environments.

In economics, the use of data that are not directly about a modeled phenomenon to reveal underlying structure in the data that do directly describe the modeled phenomena is known as model calibration. For an example of calibration applied to another phenomenon where reliance on self-reports alone leads to demonstrable bias (in this instance, values assigned by people to environmental goods), see Blackburn, Harrison and Rutström [1994] and Harrison [2006]. Examining how scores on FLAGS or other screens are conditioned on independent demographic and other variables can allow us to build up structural models of the phenomenon that incorporate hypothesized causal mechanisms. These, in turn, guide the design of experiments to supply data that can support or undermine the structural hypotheses in question.

7. Summary and Conclusions

We studied Danish adult gambling behavior with an emphasis on discovering patterns relevant to public health forecasting and economic welfare assessment of policy. The survey is a precursor to experimental work that is in turn intended to inform development of a structural model of latent vulnerability to gambling problems. By itself, the survey is primarily an indicator of prevalence that provides crucial data for recruiting informative samples for further work. This purpose, however, leads to methodological innovations in prevalence estimation, by comparison with previous efforts that have been focused on determining proportions of populations that would be recommended for clinical interventions. The methodological innovations include measurement of formative in addition to reflective constructs, estimation of prospective risk for developing GD rather than risk of being falsely negatively diagnosed, analysis with attention to sample weights and correction for sample selection bias, estimation of the impact of trigger questions on prevalence estimates and sample characteristics, and distinguishing between total and marginal effects of risk-indicating factors. The most significant novelty in our design is that nobody was excluded on the basis of their response to a trigger question about previous gambling history, allowing us to assess the impact of the more typical methodological practice.

We find that 79.7% of the realized sample are indicated for no detectable prospective risk for GD as measured by the FLAGS, 12% are indicated for Early Risk, 3.9% are indicated for Intermediate Risk, 3.3% are indicated for Advanced Risk, and 1.1% are indicated as afflicted with GD. Using sample weights and controlling for sample selection has a significant negative effect on prevalence rates, and the corrected estimates of gambling risk show that 87.6% of the population are indicated for No Detectable Risk, 5.4% are indicated for Early Risk, 1.7% are indicated for Intermediate Risk, 1.7% are indicated for Advanced Risk, and 2.6% are indicated as DGs.

There are significant (unconditional and conditional) correlations of the FLAGS prospective risk levels with the negative misdiagnosis risk levels of screens for reflective constructs (*DSM*-based and PGSI), related instruments measuring extent of other formative gambling-related constructs, an instrument measuring alcohol dependence and abuse, and instruments measuring anxiety, depression and impulsivity. All correlations are positive and statistically significantly different than zero: for example, someone that scores more highly on PGSI also scores more highly on FLAGS.

In interpreting the relationships among survey screens, attention must be given to the different methodologies that informed the designs of the instruments. The correlation reported above is telling us more than that the instruments are tracking, with somewhat different measures, the same underlying phenomena. The FLAGS was designed, through its mixture of reflective and formative constructs, to measure the risk that a given person will develop a gambling disorder, whereas the other screens are intended to measure the risk that a short screen score misses positive identifications that would be obtained in a clinical setting. It is hardly surprising that risk conceived in these two ways should be correlated. Equally, however, given the fact that the conceptions are different, it would be surprising and indeed somewhat troubling if the approaches yielded almost identical results. The relationship between observable formation of gambling disorders and frequency estimation of already existing disorders, relative to diagnostic tools, is directly relevant to design of public health intervention policies, which to date has not been pursued in the research literature.

We find significant effects of imposing a threshold gambling history on the measurement of prospective risk for GD. When 500 kroner and 40 kroner loss thresholds are imposed as filters, the consequence is a significant understatement of the fraction of Danes with detectable prospective risk levels. Hence the standard practice in surveys of using these thresholds leads to an underestimate of the prevalence of gambling problems in the general population.

Taken together, these results remind us that qualitative effects of using sample weights and correcting for sample selection bias are inherently unpredictable a priori, reinforcing the importance of eliciting responses using methods that allow such bias to be corrected. There may often be correlations between subjects' motivation to furnish full information, from which they can opt out by answering "no" to a trigger or gateway question, and the dependent variable of interest. It might be conjectured on an intuitive basis that such correlation would always underestimate prevalence because DGs might prefer to conceal their condition. However, there is also anecdotal evidence, including some from our own experience in other populations, that DGs are more interested in focusing attention on gambling than are people for whom gambling is an activity of casual and slight importance. However, among such people are the DGs of the future. To the extent that their risk can be detected before they themselves, or their families or associates, become aware of it, public health policies could have enhanced beneficial impact and the welfare consequences of changes in the regulation of gambling could be more accurately estimated and assessed.

Ultimately, these objectives will require more than survey data. The 'gold standard' for scientific knowledge, as opposed to only treatment effectiveness, is understanding of the causal structures that relate biological vulnerabilities, cultural facilitators and safeguards, cognitive capacities and information, policy responses, and clinical resources. In short, the aim is for a theory of the complete causal nexus underlying GD. This will call upon experimentation, both natural and designed, and carried out in both the laboratory and the field. Constructing the experimental agenda in turn depends on detailed and sophisticated statistical modeling of samples and of our methods of recruiting and probing them.

Table 1: Sample Tabulation of FLAGS Risk Levels with *DSM* and PGSI Categorizations

A. Full FLAGS Sample

FLAGS Risk Level	N	Percent
No Detectable Risk	6,698	79.7%
Early Risk	1,010	12.0%
Intermediate Risk	328	3.9%
Advanced Risk	274	3.3%
Problem Gambler	95	1.1%
Total	8.405	100%

B. *DSM* Sub-Sample

FLAGS Risk Level	N	Percent	*DSM* Categorizations	
			Non-Gambler	Pathological Gambler
No Detectable Risk	1,360	81.4%	1,360	0
Early Risk	177	10.6%	177	0
Intermediate Risk	66	3.9%	66	0
Advanced Risk	56	3.4%	55	1
Problem Gambler	12	0.7%	11	1
Total	1,671	100%	1,669	2

C. PGSI Sub-Sample

FLAGS Risk Level	N	Percent	PGSI Categorizations			
			Non-Gambler	Low Risk	Moderate Risk	Problem Gambler
No Detectable Risk	1,399	79.6%	1,291	93	14	1
Early Risk	229	13.0%	161	53	15	0
Intermediate Risk	62	3.5%	27	25	10	0
Advanced Risk	47	2.7%	13	12	19	3
Problem Gambler	20	1.1%	0	0	2	18
Total	1,757	100%	1,492	183	60	22

Table 2: Predicted FLAGS Risk Levels

A. No Sample Weights

FLAGS Risk Level	Prediction	95% Confidence Interval
No Detectable Risk	79.7%	79.3% ↔ 80.1%
Early Risk	12.0%	11.8% ↔ 12.1%
Intermediate Risk	3.9%	3.8% ↔ 4.0%
Advanced Risk	3.3%	3.2% ↔ 3.4%
Problem Gambler	1.1%	1.0% ↔ 1.1%

B. Sample Weights

FLAGS Risk Level	Prediction	95% Confidence Interval
No Detectable Risk	79.2%	78.9% ↔ 79.5%
Early Risk	12.1%	11.9% ↔ 12.2%
Intermediate Risk	3.6%	3.6% ↔ 3.7%
Advanced Risk	3.8%	3.7% ↔ 3.9%
Problem Gambler	1.3%	1.2% ↔ 1.3%

C. Sample Weights and Sample Selection Correction

FLAGS Risk Level	Prediction	95% Confidence Interval
No Detectable Risk	87.6%	87.3% ↔ 87.8%
Early Risk	5.4%	5.3% ↔ 5.5%
Intermediate Risk	1.7%	1.7% ↔ 1.8%
Advanced Risk	2.6%	2.6% ↔ 2.7%
Problem Gambler	2.6%	2.6% ↔ 2.7%

Note: The predicted levels of risk are based on the estimated semi-nonparametric Ordered Response model explained in the text. Predictions for blocks A and B are made for the 8,405 participants, and predictions for block C are made for the 65,580 in the sampling frame.

Figure 1: Marginal Effect of Being Female
on Probability of *FLAGS* Gambling Risk Level

Semi-Nonparametric Ordered Response model with
sample weights and sample selection corrections
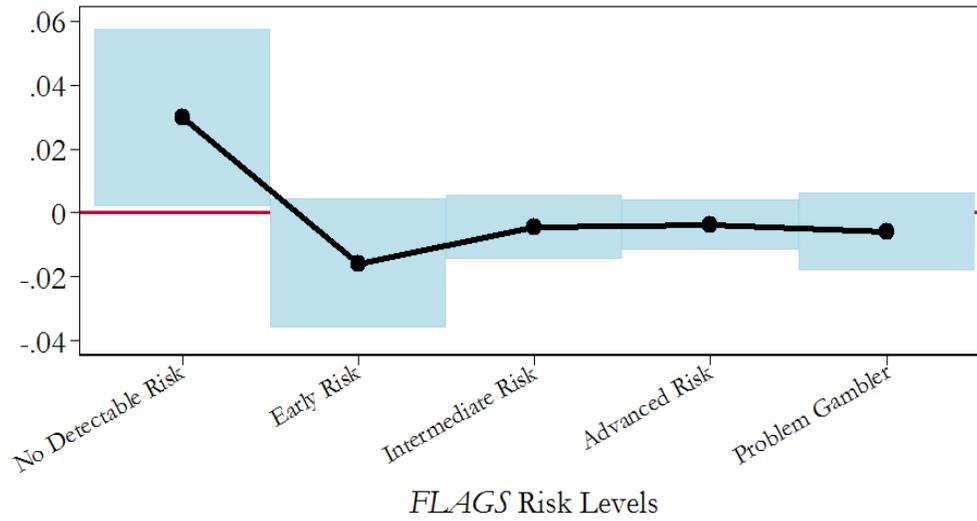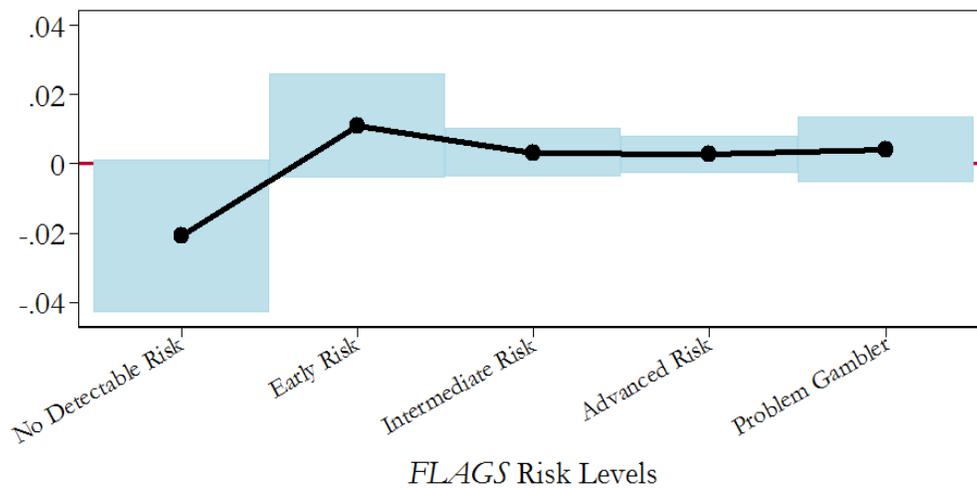Point estimate of effect and 95% confidence interval

Figure 2: Marginal Effect of Having Low Income
on Probability of *FLAGS* Gambling Risk Level

Low Income defined by income less than 300,000 kroner p.a.
Semi-Nonparametric Ordered Response model with
sample weights and sample selection corrections
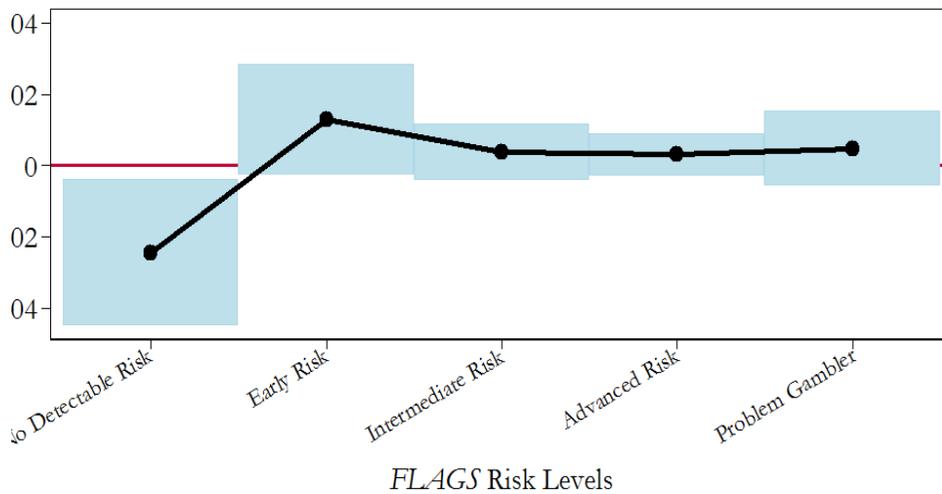Point estimate of effect and 95% confidence interval

FLAGS Risk Levels

Figure 3: Marginal Effect of FLAGS Being First
on Probability of *FLAGS* Gambling Risk Level

Semi-Nonparametric Ordered Response model with
sample weights and sample selection corrections
Point estimate of effect and 95% confidence interval

FLAGS Risk Levels

Figure 4: Marginal Effect of Randomized Questions
on Probability of *FLAGS* Gambling Risk Level

Semi-Nonparametric Ordered Response model with
sample weights and sample selection corrections
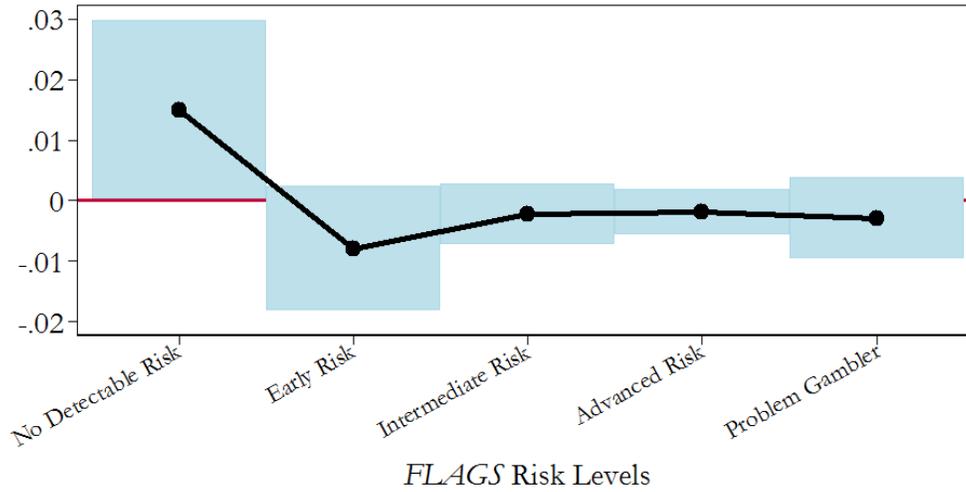Point estimate of effect and 95% confidence interval

Figure 5: Marginal Effect of Lifetime Frame
on Probability of *FLAGS* Gambling Risk Level

Semi-Nonparametric Ordered Response model with
sample weights and sample selection corrections
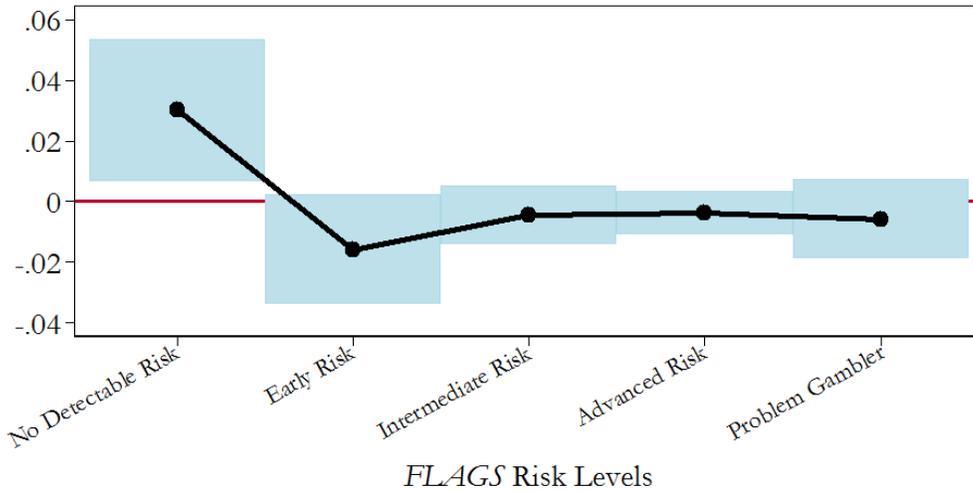Point estimate of effect and 95% confidence interval

Figure 6: Comparison of True *FLAGS* Responses and Inferred Responses if Using Gambling History Triggers
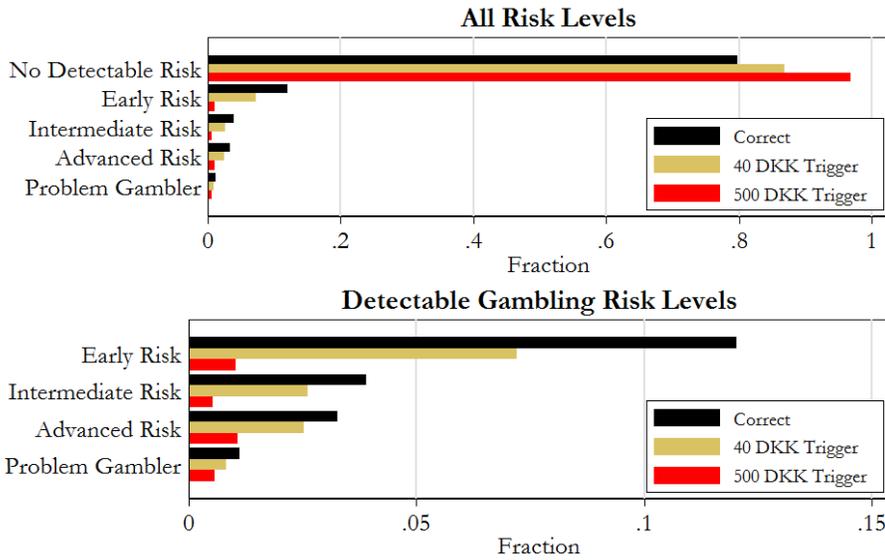
All Risk Levels

Detectable Gambling Risk Levels

Figure 7: Comparison of True *PGSI* Responses and Inferred Responses if Using Gambling History Triggers

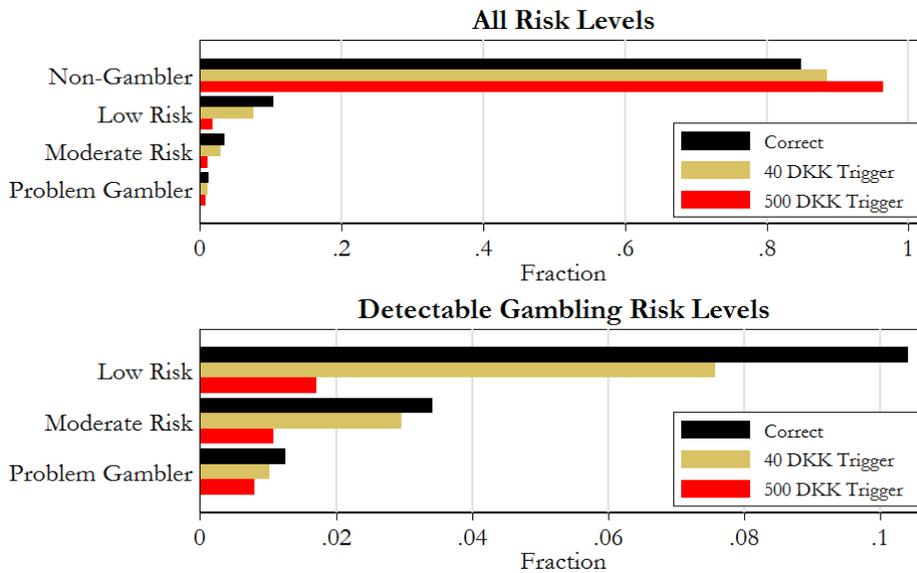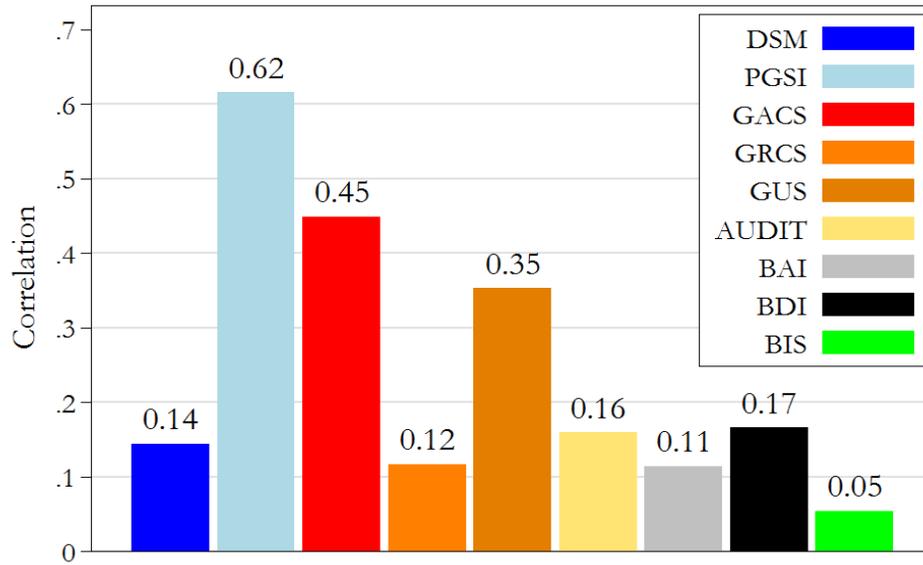**All Risk Levels**

Detectable Gambling Risk Levels

Figure 8: Correlation of *FLAGS* Scores with Other Instruments

References

Algeria, Analucia; Petry, Nancy; Hasin, Deborah; Liu, Shang-Min; Grant, Bridget, and Blanco, Carlos, "Disordered Gambling Among Racial and Ethnic Groups in the US: Results from the National Epidemiologic Survey on Alcohol and Related Conditions," *CNS Spectrums*,14, 2009, 132-142.

American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders III* (Washington, DC: American Psychiatric Association, 1980).

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

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

Babor, Thomas F.; Higgins-Biddle, John C.; Saunders, John B., and Monteiro, Maristela G., *The Use Disorders Identification Test. Guidelines for Use in Primary Care* (Geneva, CH: World Health Organization, 2001).

Beck, Aaron T.; Epstein, Norman; Brown, Gary; and Steer, Robert A., "An Inventory for Measuring Clinical Anxiety: Psychometric Properties," *Journal of Consulting and Clinical Psychology*, 56(6), 1988, 893-897.

Beck, Aaron T., and Steer, Robert A., *Manual for the Beck Anxiety Inventory* (San Antonio, TX: Psychological Corporation, 1990).

Beck, A. T.; Ward, C. H.; Mendelson, M.; Mock, J., and Erbaugh, J., "An Inventory for Measuring Depression," *Archives of General Psychiatry*, 4, 1961, 561-571.

Billi, Rosa; Stone, Christine A.; Abbott, Max, and Yeung, Kristal, "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, 2015, 274-296.

Billi, Rosa; Stone, Christine A.; Marden, P., and Yeung, Kristal, *The Victorian Gambling Study: A Longitudinal Study of Gambling and Health in Victoria, 2008–2012* (Victoria, Australia: Victorian Responsible Gambling Foundation, 2014).

Blackburn, McKinley; Harrison, Glenn W., and Rutström, E. Elisabet, "Statistical Bias Functions and Informative Hypothetical Surveys," *American Journal of Agricultural Economics*, 76, 1994, 1084-1088.

Blanco, Carlos; Hasin, Deborah. S.; Petry, Nancy; Stinson, Frederick S., and Grant, Bridget F., "Sex Differences in Subclinical and DSM-IV Pathological Gambling: Results from the National Epidemiologic Survey on Alcohol and Related Conditions," *Psychological Medicine*, 36, 2006, 943-953.

Blaszczynski, Alex, and Nower, Lia, "A Pathways Model of Problem and Pathological Gambling," *Addiction*, 97, 2002, 487-499.

Bonke, Jens, and Borregaard, Karen, *Ludomani i Danmark. Udbredelsen af pengespil og problemspillere* (København: Socialforskningsinstituttet, 2006.)

Bonke, Jens, and Borregaard, Karen, "The Prevalence of Problematic Gambling Behaviour: A Scandinavian Comparison," *Scandinavian Journal of Public Health*, 37(6), 2009, 654–660.

Breen, Robert B., and Zuckerman, Marvin, "'Chasing' in Gambling Behavior: Personality and Cognitive Determinants," *Personality and Individual Differences*, 27, 1999, 1097-1111.

Committee on the Social and Economic Impact of Pathological Gambling, *Pathological Gambling: A Critical Review* (Washington: National Academy Press, 1999).

Cox, Brian; Yu, Nancy; Afifi, Tracie D., and Ladouceur, Robert, "A National Survey of Gambling Problems in Canada," *Canadian Journal of Psychiatry*, 50(4), 2005, 213-217.

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

Ekholm, Ola; Eiberg, Stig; Davidsen, Michael; Holst, Maria; Larsen, Christina, and Juel, Knud, "The Prevalence of Problem Gambling in Denmark in 2005 and 2010: A Sociodemographic and Socioeconomic Characterization," *Journal of Gambling Studies*, 30(1), 2014, 1-10.

Ferris, Jackie, and Wynne, Harold, *The Canadian Problem Gambling Index: Final Report* (Ottawa: Canadian Centre on Substance Abuse, February 2001).

Fisher, Susan, "Measuring the Prevalence of Sector-Specific Problem Gambling: A Study of Casino Patrons," *Journal of Gambling Studies*, 16(1), 2000, 25-51.

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

Gerstein, Dean; Hoffman, John; Larison, Cindy; Engelman, Laszlo; Murphy, Sally; Palmer, Amanda; Chucro, Lucian; Tace, Marianna; Johnson, Robert; Buie, Tracy; Hill, Mary Ann; Volberg, Rachel; Harwood, Henrick; Tucker, Adam; Christiansen, Eugene; Cummings, Will, and Sinclair, Sebastian, *Gambling Impact and Behavior Study: Report to the National Gambling Impact Study Commission* (Chicago: National Opinion Research Center at the University of Chicago, 1999).

Harrison, Glenn W., "Experimental Evidence on Alternative Environmental Valuation Methods," *Environmental and Resource Economics*, 34, 2006, 125-162.

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

Analysis of Risk, Robinson College of Business, Georgia State University, 2014.

Harrison, Glenn W.; Lau, Morten I., and Ross, Don, "Psychiatric Measures of Gambling Problems in the General Population: A Reconsideration," *CEAR Working Paper 2016-09*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2016.

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

Harrison, Glenn W.; Lau, Morten I.; Rutström, Elisabet E., and Sullivan, Melonie B., "Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues," in J. Carpenter, G.W. Harrison, and J. A. List (eds.), *Field Experiments in Economics* (Greenwich, CT: JAI Press, 2005; Research in Experimental Economics, Volume 10).

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

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

Heckman, James J., and Robb, Richard, "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data* (New York: Cambridge University Press, 1985).

Hotz, V. Joseph, "Designing an Evaluation of JTPA," in C. Manski and I. Garfinkel (eds.), *Evaluating Welfare and Training Programs* (Cambridge: Harvard University Press, 1992).

Kessler, Ronald C; Abelson, Jamie; Demler, Olga; Escobar, Javier I.; Gibbon, Miriam; Guyer, Margaret E.; Howes, Mary J.; Jin, Robert; Vega, William A.; Walters, Ellen E.; Wang, Phillip; Zaslavsky, Alan, and Zheng, Hui, "Clinical Calibration of DSM-IV Diagnoses in the World Mental Health (WMH) Version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMH-CIDI)," *International Journal of Methods in Psychiatric Research*, 13(2), 2004, 122-139.

Kessler, R.C.; Hwang, I.; LaBrie, R.; Petukhova, M.; Sampson, N.A.; Winters, K.C., and Shaffer, H.J., "DSM-IV Pathological Gambling in the National Comorbidity Survey Replication," *Psychological Medicine*, 38, 2008, 1351-1360.

Levy, Paul S., and Lemeshow, Stanley, *Sampling of Populations: Methods and Applications* (New York: Wiley 2008, Fourth Edition)

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

Patton, Jim H.; Stanford, Matthew S., and Barratt, Ernest S., "Factor Structure of the Barratt Impulsiveness Scale," *Journal of Clinical Psychology*, 51(6), 1995, 768-774.

Petry, Nancy; Stinson, Frederick S., and Grant, Bridget F., "Comorbidity of DSM-IV Pathological Gambling and Other Psychiatric Disorders: Results From the National Epidemiologic Survey on Alcohol and Related Conditions," *Journal of Clinical Psychiatry*, 66, 2005, 564-574.

Petry, Nancy M.; Blanco, Carlos; Auriacombe, Marc; Borges, Guilherme; Bucholz, Kathleen; Crowley, Thomas J.; Grant, Bridget F.; Hasin, Deborah S., and O'Brien, Charles, "An Overview of and Rationale for Changes Proposed for Pathological Gambling in DSM-5," *Journal of Gambling Studies*, 30, 2014, 493-502.

Pierre, Joseph, "Overdiagnosis, Underdiagnosis, Synthesis: A Dialectic for Psychiatry and the DSM," in J. Paris and J. Phillips (eds.) *Making the DSM-5: Concepts and Controversies* (New York: Springer, 2013).

Pietrzak, R.; Morasco, B.; Blanco, C.; Grant, B., and Petry, N., "Gambling Level and Psychiatric and Medical Disorders in Older Adults: Results from the National Epidemiologic Survey on Alcohol and Related Conditions," *American Journal of Geriatric Psychiatry*, 15, 2007, 301-313.

Raylu, Namrata, and Oei, Tian P., "The Gambling Urge Scale: Development, Confirmatory Factor Validation, and Psychometric Properties," *Psychology of Addictive Behaviors*, 18(2), 2004a, 100-105.

Raylu, Namrata, and Oei, Tian P., "The Gambling Related Cognitions Scale (GRCS): Development, Confirmatory Factor Validation and Psychometric Properties," *Addiction*, 99(6), 2004b, 757-769.

Redish, A. David, "Implications of the Multiple-Vulnerabilities Theory of Addiction for Craving and Relapse," *Addiction*, 104, 1940-1941.

Redish, A. David; Jensen, Steve, and Johnson, Adam, "A Unified Framework for Addiction: Vulnerabilities in the Decision Process," *Behavioral and Brain Sciences*, 31, 415-487.

Regier, Darrel; Kaelber, Charles T.; Rae, Donald S.; Farmer, Mary E.; Knauper, Barbel; Kessler, Ronald C., and Norquist, Grayson S., "Limitations of Diagnostic Criteria and Assessment Instruments for Mental Disorders," *Archives of General Psychiatry*, 55, February 1998, 109-115.

Ross, Don; Sharp, Carla; Vuchinich, Rudolph, and Spurrett, David, *Midbrain Mutiny: The Picoeconomics and Neuroeconomics of Disordered Gambling* (Cambridge MA: MIT Press, 2008).

Sadler, John, "Considering the Economy of DSM Alternatives," in J. Paris and J. Phillips (eds.), *Making the DSM-5: Concepts and Controversies* (New York: Springer, 2013).

Schellinck, Tony; Schrans, Tracy; Bliemel, Michael, and Schellinck, Heather, *FLAGS: Development of an Instrument for Identifying Risk for Problem Gambling among Slot Machine Gamblers in Ontario* (Guelph,

ON: Ontario Problem Gambling Research Centre, 2010).

Schellinck, Tony; Schrans, Tracy; Schellinck, Heather, and Bliemel, Michael, *Raising the FLAGS: A Pilot Study Adapting FLAGS, a Next-Generation Gambling Risk Assessment Instrument, for Use in Identifying Risk Among General Gambling Populations* (Guelph, ON: Ontario Problem Gambling Research Centre, 2011).

Schellinck, Tony; Schrans, Tracy; Bliemel, Michael, and Schellinck, Heather, "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*, 2015a, 140-173.

Schellinck, Tony; Schrans, Tracy; Bliemel, Michael, and Schellinck, Heather, "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*, 2015b, 174-200.

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

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

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

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

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

Strong, David R.; Breen, Robert B.; Lesieur, Henry R., and Lejuez, C.W., "Using the Rasch Model to Evaluate the South Oaks Gambling Screen for Use with Nonpathological Gamblers," *Addictive Behaviors*, 28, 2003, 1465-1472.

Strong, David R.; Breen, Robert B., and Lejuez, C.W., "Using Item Response Theory to Examine Gambling Attitudes and Beliefs," *Personality and Individual Differences*, 36, 2004a, 1515-1529.

Strong, David R., and Kahler, C. W., "Evaluation of the Continuum of Gambling Problems using the DSM-IV," *Addiction*, 102, 5, 2007, 713-721.

Strong, David R.; Lesieur, Henry R.; Breen, Robert B.; Stinchfield, Randy, and Lejuez, C.W., "Using a Rasch Model to Examine the Utility of the South Oaks Gambling Screen Across Clinical and

Community Samples," *Addictive Behaviors*, 29, 2004b, 465-481.

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

Wardle, Heather; Sproston, Kerry; Orford, Jim; Erens, Bob; Griffiths, Mark; Constantine, Rabecca, and Pigott, Sarah, *British Gambling Prevalence Survey 2007* (London: National Centre for Social Research, 2007).

Young, Matthew M., and Wohl, Michael J., "The Gambling Craving Scale: Psychometric Validation and Behavioral Outcomes," *Psychology of Addictive Behaviors*, 23(3), 2009, 512-522.