J Gambl Stud (2018) 34:225–253
https://doi.org/10.1007/s10899-017-9707-1

ORIGINAL PAPER

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

Glenn W. Harrison[1,6,7] · **Lasse J. Jessen**[2] · **Morten I. Lau**[3,4] · **Don Ross**[5,6,7]

Published online: 13 July 2017
© Springer Science+Business Media, LLC 2017

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

---

Electronic supplementary material The online version of this article (doi:10.1007/s10899-017-9707-1) contains supplementary material, which is available to authorized users.

✉ Don Ross
don.ross931@gmail.com

Glenn W. Harrison
gharrison@gsu.edu

Lasse J. Jessen
Jessen@economics.uni-kiel.de

Morten I. Lau
mla.eco@cbs.dk

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

[2] Department of Economics, University of Kiel, Kiel, Germany

[3] Copenhagen Business School, Frederiksberg, Denmark

[4] Durham University Business School, Durham University, Durham, UK

[5] School of Sociology and Philosophy, University College Cork, Cork, Ireland

[6] School of Economics, University of Cape Town, Cape Town, South Africa

[7] Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, Atlanta, GA, USA

nobody was excluded on the basis of their response to a ‘trigger’ or ‘gateway’ question about previous gambling history. Our sample consists of 8405 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.

Keywords Gambling disorder prevalence · Denmark · Prevalence survey methodology · Formative and reflective constructs · DSM criteria · FLAGS

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 the "Classification of Disordered Gamblers" section 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 the "Methodological Issues in Research Applications of Gambling Disorder Screens" section we review the methodological limitations of standard gambling disorder screens that our Danish survey was designed to transcend. The "Sample Frame and Survey Instruments for the Danish Study" section describes our survey design. In the "Results" section we describe our results. The "Discussion" section provides a concluding critical discussion.

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 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.

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 et al. (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.

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.”

---

[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 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 is 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 1 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 1 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 et al. (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 et al. (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

---

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

(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.

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.

---

[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.

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 8405, which is 12.8% of the sample frame of 65,592 Danes contacted. Of those contacted that did not complete the survey, 3331 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 and 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 Supplementary 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. (2015a, b). 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 intercorrelation, 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.

---

[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.

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), and the other five as reflective. The complete list of statements can be found in Supplementary 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.

---

[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.

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). Supplementary 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.

---

[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.

(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 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.

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 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.

---

[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.

[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 et al. (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; Chapter 16).

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.

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 8405 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 1671 and 1757 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 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.

---

[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.

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 vs. not applied) and subcontractor (*Analyse Danmark* or *Userneeds*) are also added as control variables in the models. Supplementary Appendix E documents the

---

[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)."

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

FLAGS risk level	N	Percent (%)
A. Full FLAGS sample		
No detectable risk	6698	79.7
Early risk	1010	12.0
Intermediate risk	328	3.9
Advanced risk	274	3.3
Problem gambler	95	1.1
Total	8405	100

FLAGS risk level	N	Percent (%)	*DSM* categorizations	
			Non-gambler	Pathological gambler
B. *DSM* sub-sample				
No detectable risk	1360	81.4	1360	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	1671	100	1669	2

FLAGS risk level	N	Percent (%)	PGSI categorizations			
			Non-gambler	Low risk	Moderate risk	Problem gambler
C. PGSI sub-sample						
No detectable risk	1399	79.6	1291	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	1757	100	1492	183	60	22

detailed estimates for each model. Table 2 shows 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 panel A 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 8405 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:

Table 2 Predicted FLAGS risk levels

FLAGS risk level	Prediction	95% Confidence interval
A. No sample weights		
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		
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		
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%

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 8405 participants, and predictions for block C are made for the 65,580 in the sampling frame

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 vs. 12.1% in model B), a reduction of just over 2 percentage points in the probability of being indicated for Intermediate Risk (1.7 vs. 3.6%), a reduction of just over 1 percentage points in the probability of being indicated for Advanced Risk (2.6 vs. 3.8%), and an increase in the probability of being indicated for Disordered Gambling (2.6 vs. 1.3% in model B).

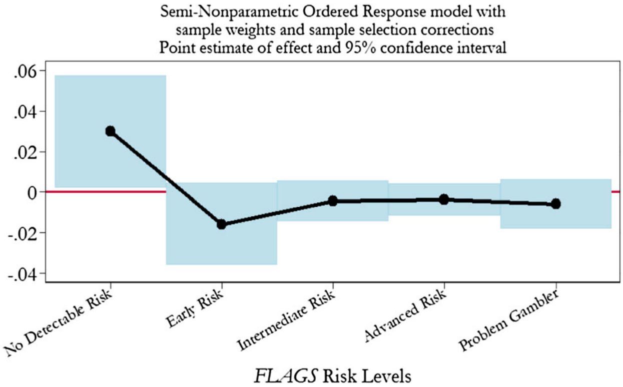


Fig. 1 Marginal effect of being female on probability of FLAGS gambling risk level

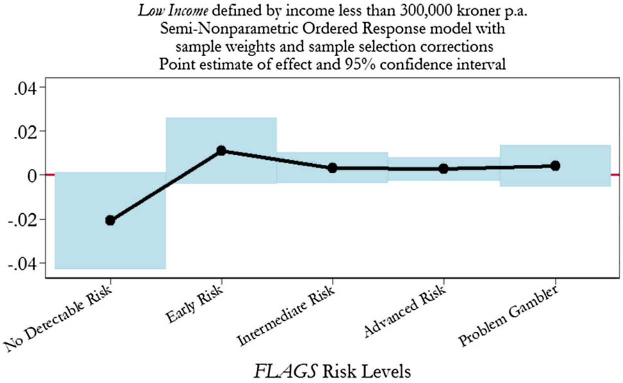


Fig. 2 Marginal effect of having low income on probability of FLAGS gambling risk level

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 Fig. 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% level.[18]

---

[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.

Comparisons

We can compare our results to previous Danish gambling prevalence studies by Bonke and Borregaard (2006, 2009) and Ekholm et al. (2014). 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 8153 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, 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. (2014) 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 5686 respondents in 2005 and 14,670 respondents in 2010. Ekholm et al. (2014) find that the prevalence rate of past-year GD is 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]

---

[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 1232 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.

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 (Fig. 3).[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 (Fig. 4).

We analyzed the use of a lifetime frame for FLAGS questions as a treatment (Fig. 5). This can be understood as interrogating the interpretation of 'risk' as prospective that was intended in the design 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

---

[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.

[22] Ekholm et al. (2014) 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. (2014; 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 χ^2 test of the hypothesis of no association has a p-value less than 0.001.

[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 χ^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.

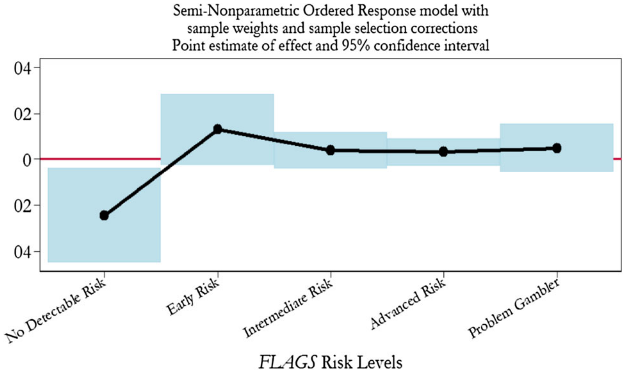


Fig. 3 Marginal effect of FLAGS being first on probability of FLAGS gambling risk level

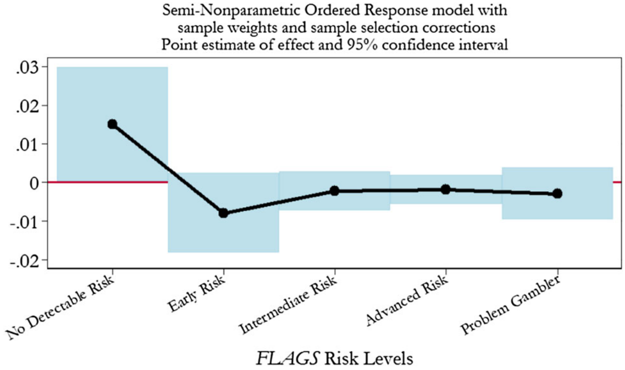


Fig. 4 Marginal effect of randomized questions on probability of FLAGS gambling risk level

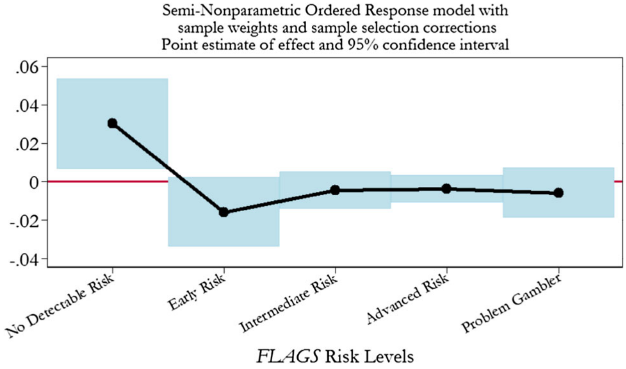


Fig. 5 Marginal effect of lifetime frame on probability of FLAGS gambling risk level

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.

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 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):

> [the] 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 Fig. 6 rescales just on the detectable gambling risks, to provide more information on these effects. We find that 95% of the sample of 8405 had not lost 500 kroner in gambling in 1 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 detectable 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, 3206 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. 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]

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.

---

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

[28] As mentioned previously, 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 Fig. 7 for the *DSM*-based screen assessment.

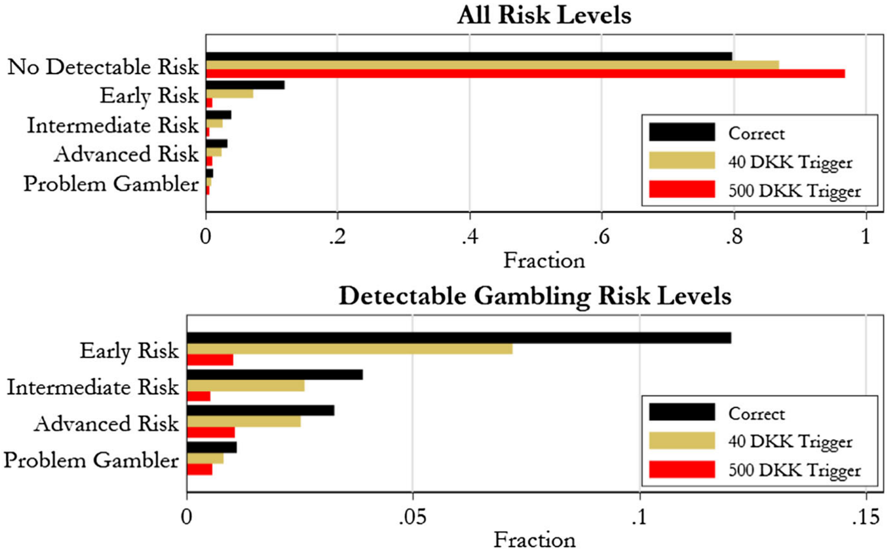


Fig. 6 Comparison of true FLAGS responses and inferred responses if using gambling history triggers

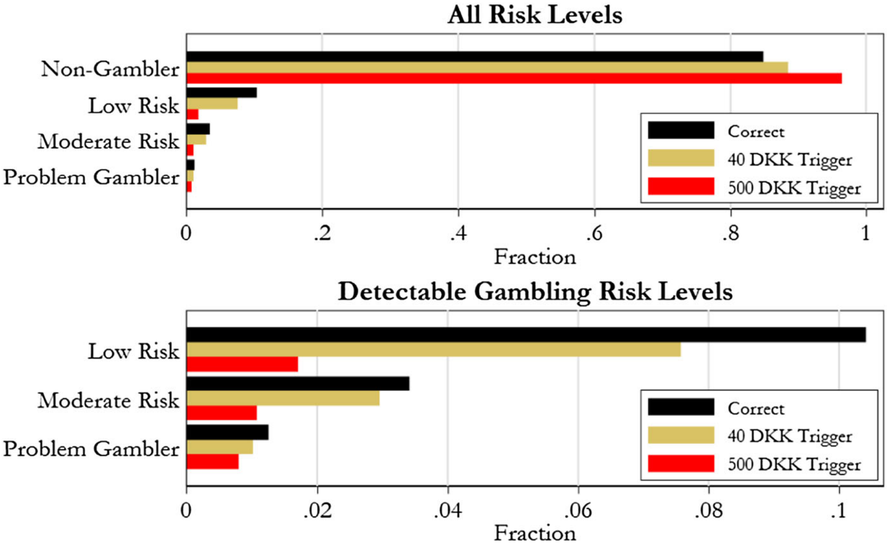


Fig. 7 Comparison of true PGSI responses and inferred responses if using gambling history triggers

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

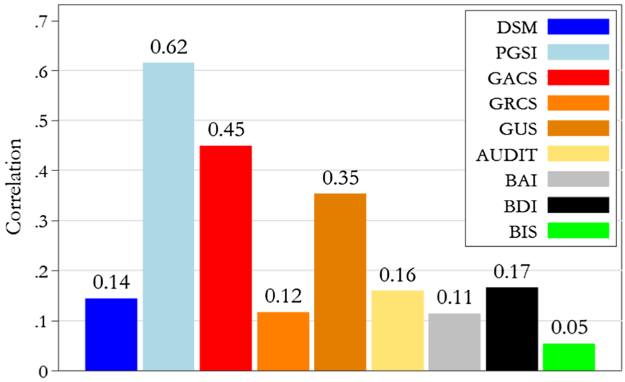


Fig. 8 Correlation of FLAGS scores with other instruments

familiar from other marginal effects considered. 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 Supplementary 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.

Discussion

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 et al. (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 7756: see Wardle et al. (2011). Harrison et al. (2016; Fig. 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 et al. (2016; Fig. 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 1057 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 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 et al. (2016; Fig. 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.

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 et al. (2009) and Harrison et al. (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.

---

[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 1759 being asked the full set of questions and having any chance of being scored as “at risk.” These deviations from the CPGI screen, and the PGSI index derived from it, were “approved by the authors of the scale” (Statistics Canada 2004; p. 19).

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 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 et al. (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.

---

[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%.

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 correlations 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.

Funding We are grateful to the Danish Social Science Research Council (Project #12-130950) for financial support.

Compliance with Ethical Standards

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

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

References

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

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

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

Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *The use disorders identification test. Guidelines for use in primary care*. Geneva, CH: World Health Organization.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*(6), 893–897.

Beck, A. T., & Steer, R. A. (1990). *Manual for the beck anxiety inventory*. San Antonio, TX: Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

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

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

Blackburn, M., Harrison, G. W., & Rutström, E. E. (1994). Statistical bias functions and informative hypothetical surveys. *American Journal of Agricultural Economics, 76,* 1084–1088.

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

Blaszczynski, A., & Nower, L. (2002). A pathways model of problem and pathological gambling. *Addiction, 97,* 487–499.

Bonke, J., & Borregaard, K. (2006). *Ludomani i Danmark. Udbredelsen af pengespil og problemspillere*. København: Socialforskningsinstituttet.

Bonke, J., & Borregaard, K. (2009). The prevalence of problematic gambling behaviour: A Scandinavian comparison. *Scandinavian Journal of Public Health, 37*(6), 654–660.

Breen, R. B., & Zuckerman, M. (1999). 'Chasing' in gambling behavior: Personality and cognitive determinants. *Personality and Individual Differences, 27,* 1097–1111.

Committee on the Social and Economic Impact of Pathological Gambling. (1999). *Pathological gambling: A critical review*. Washington, DC: National Academy Press.

Cox, B., Yu, N., Afifi, T. D., & Ladouceur, R. (2005). A national survey of gambling problems in Canada. *Canadian Journal of Psychiatry, 50*(4), 213–217.

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

Ekholm, O., Eiberg, S., Davidsen, M., Holst, M., Larsen, C., & Juel, K. (2014). The prevalence of problem gambling in Denmark in 2005 and 2010: A sociodemographic and socioeconomic characterization. *Journal of Gambling Studies, 30*(1), 1–10.

Ferris, J., & Wynne, H. (2001). *The Canadian problem gambling index: Final report*. Ottawa: Canadian Centre on Substance Abuse.

Fisher, S. (2000). Measuring the prevalence of sector-specific problem gambling: A study of casino patrons. *Journal of Gambling Studies, 16*(1), 25–51.

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

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

Harrison, G. W. (2006). Experimental evidence on alternative environmental valuation methods. *Environmental and Resource Economics, 34,* 125–162.

Harrison, G. W., Lau, M. I., & Ross, D. (2016). 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.

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

Harrison, G. W., Lau, M. I., Rutström, E. E., & Sullivan, M. B. (2005). Eliciting risk and time preferences using field experiments: Some methodological issues. In J. Carpenter, G.W. Harrison, & J.A. List (Eds.), *Field experiments in economics* (Greenwich, CT: JAI Press; Research in Experimental Economics, Volume 10).

Harrison, G. W., Lau, M., & Yoo, H. I. (2014). Risk attitudes, sample selection and attrition in a longitudinal field experiment. *CEAR Working Paper 2014-04*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.

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

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

Heckman, J. J., & Robb, R. (1985). Alternative methods for evaluating the impact of interventions. In J. Heckman & B. Singer (Eds.), *Longitudinal analysis of labor market data*. New York: Cambridge University Press.

Hotz, V. J. (1992). Designing an evaluation of JTPA. In C. Manski & I. Garfinkel (Eds.), *Evaluating welfare and training programs*. Cambridge: Harvard University Press.

Kessler, R. C., Abelson, J., Demler, O., Escobar, J. I., Gibbon, M., Guyer, M. E., et al. (2004). 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), 122–139.

Kessler, R. C., Hwang, I., LaBrie, R., Petukhova, M., Sampson, N. A., Winters, K. C., et al. (2008). DSM-IV pathological gambling in the national comorbidity survey replication. *Psychological Medicine, 38,* 1351–1360.

Levy, P. S., & Lemeshow, S. (2008). *Sampling of populations: Methods and applications* (4th ed.). New York: Wiley.

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

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. *Journal of Clinical Psychology, 51*(6), 768–774.

Petry, N. M., Blanco, C., Auriacombe, M., Borges, G., Bucholz, K., Crowley, T. J., et al. (2014). An overview of and rationale for changes proposed for pathological gambling in DSM-5. *Journal of Gambling Studies, 30,* 493–502.

Petry, N., Stinson, F. S., & Grant, B. F. (2005). 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,* 564–574.

Pierre, J. (2013). Overdiagnosis, underdiagnosis, synthesis: A dialectic for psychiatry and the DSM. In J. Paris & J. Phillips (Eds.), *Making the DSM-5: Concepts and controversies*. New York: Springer.

Pietrzak, R., Morasco, B., Blanco, C., Grant, B., & Petry, N. (2007). Gambling level and psychiatric and medical disorders in older adults: Results from the national epidemiologic survey on alcohol and related conditions. *American Journal of Geriatric Psychiatry, 15,* 301–313.

Raylu, N., & Oei, T. P. (2004a). The gambling urge scale: Development, confirmatory factor validation, and psychometric properties. *Psychology of Addictive Behaviors, 18*(2), 100–105.

Raylu, N., & Oei, T. P. (2004b). The gambling related cognitions scale (GRCS): Development, confirmatory factor validation and psychometric properties. *Addiction, 99*(6), 757–769.

Redish, A. D. (2009). Implications of the multiple-vulnerabilities theory of addiction for craving and relapse. *Addiction, 104,* 1940–1941.

Redish, A. D., Jensen, S., & Johnson, A. (2008). A unified framework for addiction: Vulnerabilities in the decision process. *Behavioral and Brain Sciences, 31,* 415–487.

Regier, D., Kaelber, C. T., Rae, D. S., Farmer, M. E., Knauper, B., Kessler, R. C., et al. (1998). Limitations of diagnostic criteria and assessment instruments for mental disorders. *Archives of General Psychiatry, 55,* 109–115.

Ross, D., Sharp, C., Vuchinich, R., & Spurrett, D. (2008). *Midbrain mutiny: The picoeconomics and neuroeconomics of disordered gambling*. Cambridge, MA: MIT Press.

Sadler, J. (2013). Considering the economy of DSM alternatives. In J. Paris & J. Phillips (Eds.), *Making the DSM-5: Concepts and controversies*. New York: Springer.

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

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

Schellinck, T., Schrans, T., Schellinck, H., & Bliemel, M. (2011). *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.

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

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

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

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

Strong, D. R., Breen, R. B., & Lejuez, C. W. (2004a). Using item response theory to examine gambling attitudes and beliefs. *Personality and Individual Differences, 36,* 1515–1529.

Strong, D. R., Breen, R. B., Lesieur, H. R., & Lejuez, C. W. (2003). Using the rasch model to evaluate the south oaks gambling screen for use with nonpathological gamblers. *Addictive Behaviors, 28,* 1465–1472.

Strong, D. R., & Kahler, C. W. (2007). Evaluation of the continuum of gambling problems using the DSM-IV. *Addiction, 102*(5), 713–721.

Strong, D. R., Lesieur, H. R., Breen, R. B., Stinchfield, R., & Lejuez, C. W. (2004b). Using a rasch model to examine the utility of the south oaks gambling screen across clinical and community samples. *Addictive Behaviors, 29,* 465–481.

Wardle, H., Moody, A., Spence, S., Orford, J., Volberg, R., Jotangia, D., et al. (2011). *British gambling prevalence survey 2010.* London: Stationery Office.

Young, M. M., & Wohl, M. J. (2009). The gambling craving scale: Psychometric validation and behavioral outcomes. *Psychology of Addictive Behaviors, 23*(3), 512–522.

Appendix A: Survey Instruments

The instruments listed below were employed for use in our surveys. The format displayed below is the easiest to document the blocks and response formats, but is not literally the text given to respondents. For instance, the block headings with the FLAGS instrument were not given to respondents, but make it easier to see which questions go together.

A. Focal Adult Gambling Screen (FLAGS)

Please read the following list of statements about gambling, and for each please indicate whether or not this statement is true for you. A simple YES or NO response is all you need to provide. If some statement does not apply to you, just respond NO. The statements are in no particular order. Think about your lifetime gambling experiences [OR: your gambling experiences in the last year].

Risky Cognition: Belief (RCB)

RCB01	People who are good at gambling get more respect from others.
RCB02	Gambling is an easy way to get extra money when you need it.
RCB03	Using a system or a strategy when you play games like lottery draws, roulette, slots or other gambling machines improves your chances of winning.
RCB04	After you have been gambling and losing for a while your chances of winning improve.
RCB05	If you are on a winning streak it makes sense to keep gambling to take advantage of your luck.
RCB06	Once someone has been gambling and losing for a while they should keep playing so they don't miss out on the chance to win back their money.
RCB07	Some people are luckier than others and so they have a better chance of winning when they gamble.
RCB08	I had a big win when I first started to gamble.
RCB09	I used to win a lot when I gambled but it seems like the odds aren't as good anymore.

Risky Cognition: Motives (RCM)

RCM01	I find things a lot more interesting when I have a bet or wager riding on the outcome.
RCM02	I sometimes gamble when I am feeling down or depressed.
RCM03	I gamble when I want money.
RCM04	Even if I don't have much money I gamble to try to get a big win.
RCM05	Gambling helps me to fit in better with others.
RCM06	I want people to think I am good at gambling.
RCM07	I gamble because it feels good to beat others at the game.

Preoccupation: Desire (POD)

POD01	I would like to gamble almost every day.
POD02	Compared to many other things I can do I would rather gamble.
POD03	I wish I could gamble more often.

Impaired Control: Continue (ICC)

ICC01 Even when I only wanted to spend a few dollars gambling, I often ended up spending much more.
ICC02 Once I start gambling I sometimes find it difficult to stop.
ICC03 When I gambled I usually exceeded the amount of money I intended to spend.
ICC04 I often had trouble stopping/quitting play when I was ahead.

Risky Practices: Earlier (RBE)

RBE01 When I won big I usually kept playing rather than stop.
RBE02 I usually played at maximum bet or bet the most I could afford when I was on a winning streak.
RBE03 I often spend more time gambling than I intended.
RBE04 I am making bigger bets than I used to.
RBE05 I gamble or make bets when I get a good luck sign or I am feeling lucky.
RBE06 When I am gambling I am interested in the game, not socializing with others.

Risky Practices: Later (RBL)

RBL01 I sometimes borrowed money from others so I could go and gamble.
RBL02 I often used my credit card to get more money so I could keep gambling.
RBL03 I sometimes exceeded the amount of money I intended to spend in order to win back money I had lost.
RBL04 I gambled at work or when I was supposed to be doing something else.
RBL05 I have a strategy to improve my luck when I gamble.
RBL06 I have gambled using money that I was supposed to spend on other things.
RBL07 When gambling I usually played fast or made as many bets as quickly as I could.
RBL08 I sometimes bet when I believe I can predict the outcome of an event.

Impaired Control: Begin (ICB)

ICB01 If I have the opportunity to gamble I can’t stop myself from taking it.
ICB02 There have been times I started to gamble despite my desire not to.
ICB03 I tried unsuccessfully to stop or reduce my gambling.
ICB04 I gamble more often than I want to.

Preoccupation: Obsession (POO)

POO01 I plan my life around my gambling.
POO02 I am obsessed with gambling thinking about when I will next gamble all the time.
POO03 I spend a lot of my time thinking about gambling or how to get money to gamble.
POO04 Thoughts of gambling or playing the games fill my mind day and night.

Negative Consequences (NGC)

NGC01 When I finished gambling, I sometimes did not have enough money for parking, food, a ride home or other things I was supposed to buy for myself.
NGC02 Sometimes I had to juggle money and bills to cover the cost of my gambling.
NGC03 My performance at work was negatively affected by my gambling.
NGC04 My goals in life were jeopardized by my gambling.
NGC05 I sometimes had trouble sleeping thinking about gambling.

NGC06	I missed social events with my friends and family because I was gambling.
NGC07	I sometimes neglected family or friends in order to gamble.
NGC08	My gambling caused problems for my relationship with my spouse or partner.
NGC09	I have become somewhat of a loner because of my gambling.
NGC10	My gambling caused me to have a falling out with people I used to hang out with.
NGC11	Sometimes I felt depressed over my gambling.
NGC12	I regret that I gambled as much as I did.
NGC13	I have lied to others about my gambling.

Persistence (PST)

PST01	Regardless of negative consequences, I kept gambling whenever I had the opportunity.
PST02	I have continued to gamble for some time despite the negative way gambling has affected my life.
PST03	I kept gambling last year even though I knew it was causing major problems for me.

B. The Problem Gambling Severity Index (PGSI)

[It is easy to confuse the acronyms of the PGSI and the CPGI. The PGSI is a scored 9-item component of the CPGI, which contains 12 items]

Some of the next 9 questions may not apply to you, but please try to be as accurate as possible. Think about your lifetime gambling experiences [OR: your gambling experiences in the last year].

Answer possibilities:
0 = Never
-1 = Sometimes
-2 = Most of the time
-3 = Almost always
-9 = Don't know

1. Have you bet more than you could really afford to lose?
2. Have you needed to gamble with larger amounts of money to get the same feeling of excitement?
3. When you gambled, did you go back on another day to try to win back the money you lost?
4. Have you borrowed money or sold anything to get money to gamble?
5. Have you felt that you might have a problem with gambling?
6. Has gambling caused you any health problems, including stress or anxiety?
7. Have people criticized your betting or told you that you had a gambling problem, regardless of whether or not you thought it was true?
8. Has your gambling caused any financial problems for you or your household?
9. Have you felt guilty about the way you gamble or what happens when you gamble?

C. The Diagnostic & Statistical Manual of Mental Disorders – Edition IV (DSM-IV)

Some of the next 10 questions may not apply to you, but please try to be as accurate as possible. Think about your lifetime gambling experiences [OR: your gambling experiences in the last year].

Answer possibilities:
0 = No
-1 = Once or twice
-2 = Sometimes
-3 = Often
-4 = Don't know

1. Have you found yourself thinking about gambling (e.g. reliving past gambling experiences, planning the next time you will play or thinking of ways to get money for gambling)?
2. Have you needed to gamble with more and more money to get the amount of excitement you are looking for?
3. Have you become restless or irritable when trying to cut down or stop gambling?
4. Have you gambled to escape from problems or when you are feeling depressed, anxious or bad about yourself?
5. After losing money gambling, have you returned another day in order to get even?
6. Have you lied to your family, or others, to hide the extent of your gambling?
7. Have you made repeated unsuccessful attempts to control, cut back or stop gambling?
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?

D. The Gambling Craving Scale (GACS)

Please indicate how much you agree or disagree with the statements.

Answer possibilities:
1 = strongly disagree
2 = moderately disagree
3 = mildly disagree
4 = neither agree or disagree
5 = mildly agree
6 = moderately agree
7 = strongly agree

1 Gambling would be fun right now.
2 If I had an opportunity to gamble right now, I probably would take it.
3 If I were offered an opportunity to gamble right now, I would gamble.
4 If it were possible, I probably would gamble now.
5 I would not enjoy gambling right now.
6 Gambling would be very satisfying now.
7 Gambling now would make things seem just perfect.
8 All I want right now is to gamble.
9 I crave gambling right now.
10 My desire to gamble seems overpowering.
11 I need to gamble now.
12 I have an urge to gamble.
13 I would do almost anything to gamble now.
14 Nothing would be better than gambling now.
15 If I were gambling now, I could think more clearly.
16 I could control things better right now if I could gamble.
17 Gambling would make me less depressed.
18 I would be less irritable right now if I could gamble.

E. The Gambling Related Cognition Scale (GRCS)

Please indicate how much you agree or disagree with the statements.

Answer possibilities:
1 = strongly disagree
2 = moderately disagree
3 = mildly disagree
4 = neither agree or disagree
5 = mildly agree
6 = moderately agree
7 = strongly agree

1 Gambling makes me happier.
2 I can't function without gambling.
3 Praying helps me win.
4 Losses when gambling, are bound to be followed by a series of wins.
5 Relating my winnings to my skill and ability makes me continue gambling.
6 Gambling makes things seem better.
7 It is difficult to stop gambling as I am so out of control.
8 Specific numbers and colors can help increase my chances of winning.
9 A series of losses will provide me with a learning experience that will help me win later.
10 Relating my losses to bad luck and bad circumstances makes me continue gambling.
11 Gambling makes the future brighter.
12 My desire to gamble is so overpowering.
13 I collect specific objects that help increase my chances of winning.

14 When I have a win once, I will definitely win again.
15 Relating my losses to probability makes me continue gambling.
16 Having a gamble helps reduce tension and stress.
17 I'm not strong enough to stop gambling.
18 I have specific rituals and behaviors that increase my chances of winning.
19 There are times that I feel lucky and thus, gamble those times only.
20 Remembering how much money I won last time makes me continue gambling.
21 I will never be able to stop gambling.
22 I have some control over predicting my gambling wins.
23 If I keep changing my numbers, I have less chances of winning than if I keep the same numbers every time.

F. The Gambling Urge Screen (GUS)

The next 8 questions have to do with your feelings and thoughts right now. Please indicate how much you agree or disagree with the statements.

Answer possibilities:
1 = strongly disagree
2 = moderately disagree
3 = mildly disagree
4 = neither agree or disagree
5 = mildly agree
6 = moderately agree
7 = strongly agree

1 All I want to do now is to gamble.
2 It would be difficult to turn down a gamble this minute.
3 Having a gamble now would make things seem just perfect.
4 I want to gamble so bad that I can almost feel it.
5 Nothing would be better than having a gamble right now.
6 I crave a gamble right now.
7 I don't need to have a gamble now.
8 If I had the chance to have a gamble, I don't think I would gamble.

G. The Alcohol Use Disorders Identification Test (AUDIT)

Now I am going to ask you some questions about your use of alcoholic beverages during the past year. Because alcohol use can affect many areas of health, and may interfere with certain medications, it is important for us to know how much you usually drink and whether you have experienced any problems with your drinking. Please try to be as honest and as accurate as you can be.

1 How often do you have a drink containing alcohol?
0 - Never
1 - Monthly or less
2 - 2-4 times a month
3 - 2-3 times a week
4 - 4 or more times a week

2 How many drinks containing alcohol do you have on a typical day when you are drinking?
0 - 1 or 2
1 - 3 or 4
2 - 5 or 6
3 - 7 to 9
4 - 10 or more

3 How often do you have six or more drinks on one occasion?
0 - Never
1 - Less than monthly
2 - Monthly
3 - Weekly
4 - Daily or almost daily

4 How often during the last year have you found that you were not able to stop drinking once you had started?
0 - Never
1 - Less than monthly
2 - Monthly
3 - Weekly
4 - Daily or almost daily

5 How often during the last year have you failed to do what was normally expected of you because of drinking?
0 - Never
1 - Less than monthly
2 - Monthly
3 - Weekly
4 - Daily or almost daily

6 How often during the last year have you needed a first drink in the morning to get yourself going after a heavy drinking session?
0 - Never
1 - Less than monthly
2 - Monthly
3 - Weekly
4 - Daily or almost daily

7 How often during the last year have you had a feeling of guilt or remorse after drinking?
0 - Never
1 - Less than monthly
2 - Monthly
3 - Weekly
4 - Daily or almost daily

8 How often during the last year have you been unable to remember what happened the night before because of your drinking?
0 - Never
1 - Less than monthly
2 - Monthly
3 - Weekly
4 - Daily or almost daily

9 Have you or someone else been injured because of your drinking?
0 - No
2 - Yes, but not in the last year
4 - Yes, during the last year

10 Has a relative, friend, doctor, or other health care worker been concerned about your drinking or suggested you cut down?
0 - No
2 - Yes, but not in the last year
4 - Yes, during the last year

H. Beck Anxiety Index (BAI)

Over their entire life, some people have had a time in their life when they were a "worrier," in the sense that they worried a lot more about things than other with the same problems. Some people have also had a time when they were much more nervous or anxious than most other people with the same problems. And some people have had a time lasting 6 months or longer when they were anxious and worried most days. The next block of statements are common symptoms of people having these experiences. Please indicate how much you have been bothered by that symptom during the worst of these experiences in your life.

Answer possibilities:
1 = Not at all
2 = Mildly but it didn't bother me much.
3 = Moderately - it wasn't pleasant at times.
4 = Severely - it bothered me a lot.

1 Numbness or tingling
2 Feeling hot
3 Wobbliness in legs

4 Unable to relax
5 Fear of worst happening
6 Dizzy or lightheaded
7 Heart pounding/racing
8 Unsteady
9 Terrified or afraid
10 Nervous
11 Feeling of choking
12 Hands trembling
13 Shaky / unsteady
14 Fear of losing control
15 Difficulty in breathing
16 Fear of dying
17 Scared
18 Indigestion
19 Faint / lightheaded
20 Face flushed
21 Hot/cold sweats

I. Beck Depression Inventory (BDI)

Over their entire life, some people have had a time when they felt sad, blue, depressed or down most of the time for at least 2 weeks. Some people have also had a time over their entire life, lasting 2 weeks or more, when they didn't care about the things they usually cared about, or when they didn't enjoy the things they usually enjoyed. The next block of statements are about experiences that you might have had during one of those times, when your mood was the lowest, or you enjoyed or cared the least about things. Please pick out the one statement that best describes the way you have been feeling during that period. If none of the statements apply, just select the first one.

1 Sadness
    0 - I do not feel sad.
    1 - I feel sad much of the time.
    2 - I am sad all the time.
    3 - I am so sad or unhappy that I can't stand it.

2 Pessimism
    0 - I am not discouraged about my future.
    1 - I feel more discouraged about my future than I used to be.
    2 - I do not expect things to work out for me.
    3 - I feel my future is hopeless and will only get worse.

3 Past Failure
    0 - I do not feel like a failure.
    1 - I have failed more than I should have.
    2 - As I look back, I see a lot of failures.

3 - I feel I am a total failure as a person.

4 Loss of Pleasure
0 - I get as much pleasure as I ever did from the things I enjoy.
1 - I don't enjoy things as much as I used to.
2 - I get very little pleasure from the things I used to enjoy.
3 - I can't get any pleasure from the things I used to enjoy.

5 Guilty Feelings
0 - I don't feel particularly guilty.
1 - I feel guilty over many things I have done or should have done.
2 - I feel quite guilty most of the time.
3 - I feel guilty all of the time.

6 Punishment Feelings
0 - I don't feel I am being punished.
1 - I feel I may be punished.
2 - I expect to be punished.
3 - I feel I am being punished.

7 Self-Dislike
0 - I feel the same about myself as ever.
1 - I have lost confidence in myself.
2 - I am disappointed in myself.
3 - I dislike myself.

8 Self-Criticalness
0 - I don't criticize or blame myself more than usual.
1 - I am more critical of myself than I used to be.
2 - I criticize myself for all of my faults.
3 - I blame myself for everything bad that happens.

9 Suicidal Thoughts or Wishes
0 - I don't have any thoughts of killing myself.
1 - I have thoughts of killing myself, but I would not carry them out.
2 - I would like to kill myself.
3 - I would kill myself if I had the chance.

10 Crying
0 - I don't cry any more than I used to.
1 - I cry more than I used to.
2 - I cry over every little thing.
3 - I feel like crying, but I can't.

11 Agitation
0 - I am no more restless or wound up than usual.

1 - I feel more restless or wound up than ususal.
2 - I am so restless or agitated that it's hard to stay still.
3 - I am so restless or agitated that have to keep moving or doing something.

12 Loss of Interest
0 - I have not lost interest in other people or activities.
1 - I am less interested in other people or things than before.
2 - I have lost most of my interest in other people.
3 - It's hard to get interested in anything.

13 Indecisiveness
0 - I make decisions about as well as ever.
1 - I find it more difficult to make decisions than usual.
2 - I have much greater difficulty in making decisions than I used to.
3 - I have trouble making any decisions.

14 Worthlessness
0 - I do not feel I am worthless.
1 - I don't consider myself as worthwhile and useful as I used to.
2 - I feel more worthless as compared to other people.
3 - I feel utterly worthless.

15 Loss of Energy
0 - I have as much energy as ever.
1 - I have less energy than I used to have.
2 - I don't have enough energy to do very much.
3 - I don't have enough energy to do anything.

16 Changes in Sleeping Pattern
0 - I have not experienced any change in my sleeping pattern.
1a - I sleep somewhat more than usual.
1b - I sleep somewhat less than usual.
2a - I sleep a lot more than usual.
2b - I sleep a lot less than usual.
3a - I sleep most of the day.
3b - I wake up 1-2 hours early and can't get back to sleep.

17 Irritability
0 - I am no more irritable than usual.
1 - I am more irritable than usual.
2 - I am much more irritable than usual.
3 - I am irritable all the time.

18 Changes in Appetite
0 - I have not experienced any change in my appetite.
1a - My appetite is somewhat less than usual.

1b - My appetite is somewhat greater than usual.
2a - My appetite is much less than before.
2b - My appetite is much greater than before.
3a - I have no appetite at all.
3b - I crave food all the time.

19 Concentration Difficulty
0 - I can concentrate as well as ever.
1 - I can't concentrate as well as usual.
2 - It's hard to keep my mind on anything for very long.
3 - I find I can't concentrate on anything.

20 Tiredness or Fatigue
0 - I am no more tired or fatigued than usual.
1 - I get more tired or fatigued more easily than usual.
2 - I am too tired or fatigued to do a lot of the things I used to do.
3 - I am too tired or fatigued to do most of the things I used to do.

21 Loss of Interest in Sex
0 - I have not noticed any recent change in my interest in sex.
1 - I am less interested in sex than I used to be.
2 - I am much less interested in sex now.
3 - I have lost interest in sex completely.

J. Barratt Impulsivity Scale (BIS)

People differ in the ways they act and think in different situations. The next block of questions are designed to measure some of the ways in which you act and think. Read each statement and select the best answer.

Answer possibilities:
1 = Rarely / Never
2 = Occasionally.
3 = Often
4 = Almost Always / Always

1 I plan tasks carefully.
2 I do things without thinking.
3 I make-up my mind quickly.
4 I am happy-go-lucky.
5 I don't "pay attention."
6 I have "racing" thoughts.
7 I plan trips well ahead of time.
8 I am self controlled.
9 I concentrate easily.

10 I save regularly.
11 I “squirm” at plays or lectures.
12 I am a careful thinker.
13 I plan for job security.
14 I say things without thinking.
15 I like to think about complex problems.
16 I change jobs.
17 I act “on impulse.”
18 I get easily bored when solving thought problems.
19 I act on the spur of the moment.
20 I am a steady thinker.
21 I change residences.
22 I buy things on impulse.
23 I can only think about one thing at a time.
24 I change hobbies.
25 I spend or charge more than I earn.
26 I often have extraneous thoughts when thinking.
27 I am more interested in the present than the future.
28 I am restless at the theater or lectures.
29 I like puzzles.
30 I am future oriented.

K. Additional Questions

1. Have you experienced the death of an immediate family member (partner, child, parent or sibling) in the past 12 months? (Yes/No)

2. Have you been hospitalized for a major medical problem during the past 12 months? (Yes/No)

3a. In the past 12 months have you lost more than 40 kroner on gambling in a single day? (Yes/No)
3b. Have you ever lost more than 40 kroner on gambling in a single day? (Yes/No)

4a. In the past 12 months have you lost more than 500 kroner on gambling in a single day? (Yes/No)
4b. Have you ever lost more than 500 kroner on gambling in a single day? (Yes/No)

5. Do you currently smoke cigarettes?
6. If yes, how much do you smoke in one day?

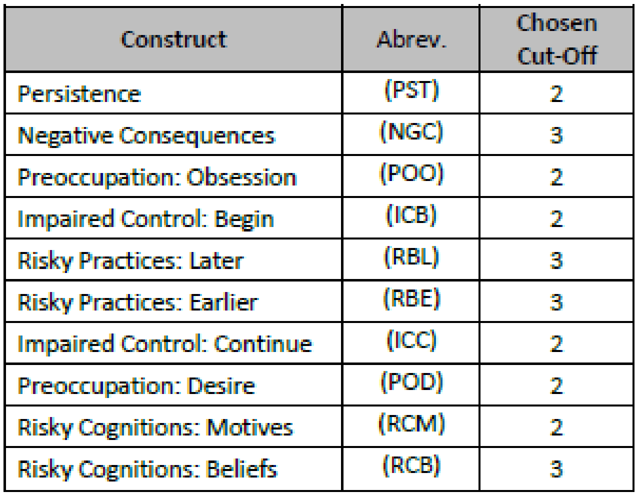
Thank you very much for your participation in the survey. Your answers are valuable to us.

Appendix B: Scoring the Survey Instruments

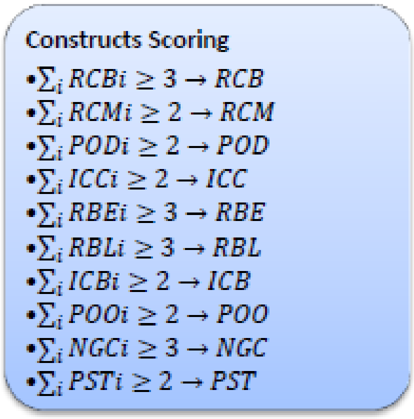
In this appendix we document how we have scored each instrument.

A. FLAGS

We have 8,422 valid survey respondents that completed the survey, and of course everyone did FLAGS. The FLAGS documentation in Schellinck et al. [2011] explains the scoring:

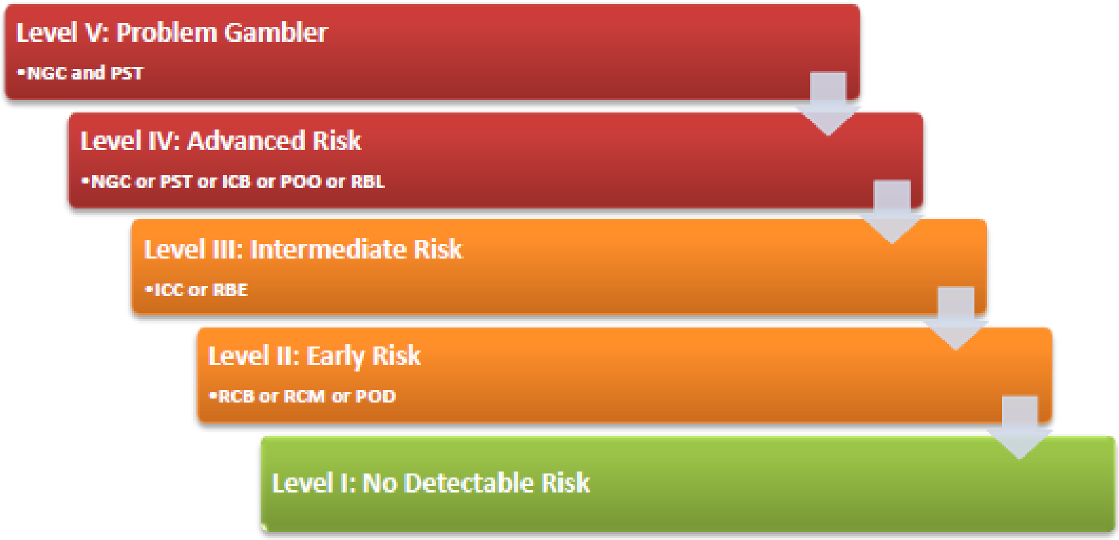


Construct	Abrev.	Chosen Cut-Off
Persistence	(PST)	2
Negative Consequences	(NGC)	3
Preoccupation: Obsession	(POO)	2
Impaired Control: Begin	(ICB)	2
Risky Practices: Later	(RBL)	3
Risky Practices: Earlier	(RBE)	3
Impaired Control: Continue	(ICC)	2
Preoccupation: Desire	(POD)	2
Risky Cognitions: Motives	(RCM)	2
Risky Cognitions: Beliefs	(RCB)	3



Constructs Scoring

- $\sum_i RCBi \geq 3 \rightarrow RCB$
- $\sum_i RCMi \geq 2 \rightarrow RCM$
- $\sum_i PODi \geq 2 \rightarrow POD$
- $\sum_i ICCi \geq 2 \rightarrow ICC$
- $\sum_i RBEi \geq 3 \rightarrow RBE$
- $\sum_i RBLi \geq 3 \rightarrow RBL$
- $\sum_i ICBi \geq 2 \rightarrow ICB$
- $\sum_i POOi \geq 2 \rightarrow POO$
- $\sum_i NGCi \geq 3 \rightarrow NGC$
- $\sum_i PSTi \geq 2 \rightarrow PST$


Level V: Problem Gambler
•NGC and PST
Level IV: Advanced Risk
•NGC or PST or ICB or POO or RBL
Level III: Intermediate Risk
•ICC or RBE
Level II: Early Risk
•RCB or RCM or POD
Level I: No Detectable Risk

Here is the raw tabulation:

```
  FLAGS Risk Level |      Freq.     Percent        Cum.
-------------------+-----------------------------------
No Detectable Risk |      6,698       79.69       79.69
        Early Risk |      1,010       12.02       91.71
 Intermediate Risk |        328        3.90       95.61
     Advanced Risk |        274        3.26       98.87
   Problem Gambler |         95        1.13      100.00
-------------------+-----------------------------------
             Total |      8,405      100.00
```

B. PGSI

We have 1,757 valid survey respondents that completed the PGSI survey questions, and of course everyone in that 1,757 also did FLAGS.

The PGSI is directly scored from the responses, once we conservatively assign a Don't Know response as a No. The classification is then to assign anyone with a score of 1 or 2 to Low Risk, a score between 3 and 7 to Moderate Risk, and a score of 8 or more to Problem Gambler.

Here is the raw tabulation:

```
PGSI Risk Level |      Freq.     Percent        Cum.
----------------+-----------------------------------
    Non-Gambler |      1,492       84.92       84.92
       Low Risk |        183       10.42       95.33
  Moderate Risk |         60        3.41       98.75
Problem Gambler |         22        1.25      100.00
----------------+-----------------------------------
          Total |      1,757      100.00
```

```
                |  Lifetime prevalence
PGSI Risk Level |         0          1 |     Total
----------------+----------------------+----------
    Non-Gambler |       650        842 |     1,492 
       Low Risk |        79        104 |       183 
  Moderate Risk |        29         31 |        60 
Problem Gambler |        11         11 |        22 
----------------+----------------------+----------
          Total |       769        988 |     1,757 

          Pearson chi2(3) =   0.9067   Pr = 0.824
 likelihood-ratio chi2(3) =   0.9009   Pr = 0.825
                    gamma =  -0.0303  ASE = 0.065
          Kendall's tau-b =  -0.0110  ASE = 0.024
```

C. DSM

We have 1,671 valid survey respondents that completed the DSM-IV survey questions, and of course everyone in that 1,671 also did FLAGS.

One initial question is how we score these responses. Many of the applications of DSM just ask YES or NO questions (apart from a DON'T KNOW). But we asked the extended list:

```
     Have you |
        found |
     yourself |
     thinking |
        about |
     gambling |
        (e.g. |
reliving past |
 gambling exp |      Freq.     Percent        Cum.
--------------+-----------------------------------
   Don't know |         13        0.78        0.78
        Often |         20        1.20        1.97
    Sometimes |         98        5.86        7.84
Once or twice |        101        6.04       13.88
           No |      1,439       86.12      100.00
--------------+-----------------------------------
        Total |      1,671      100.00
```

```
   Have you |
      found |
   yourself |
   thinking |
      about |
   gambling |
      (e.g. |
   reliving |
       past |
   gambling |
        exp |      Freq.     Percent        Cum.
------------+-----------------------------------
         -4 |         13        0.78        0.78
         -3 |         20        1.20        1.97
         -2 |         98        5.86        7.84
         -1 |        101        6.04       13.88
          0 |      1,439       86.12      100.00
------------+-----------------------------------
      Total |      1,671      100.00
```

This comes from the MR (Multiple Response) version of the DSM screen developed by Susan Fisher, “Measuring the Prevalence of Sector-Specific Problem Gambling: A Study of Casino Patrons,” *Journal of Gambling Studies*, 16(1), 2000, 25-51. She argues these are easier for respondents to follow in a non-clinical setting where one is not face-to-face with the subject. She has also used a closely-related MR version in Sue Fisher, “Developing the DSM-IV-DSM-IV Criteria to Identify Adolescent Problem Gambling in Non-Clinical Populations,” *Journal of Gambling Studies*, 16(2/3), 2000, 253-273. A MR version was also used in the British Gambling Prevalence Study of 2007 (p. 135) and 2010 (p. 154), who also proposed a “continuous” scoring for each question rather than the mapping from MR categories to a “binary” YES or NO.

We follow the first Fisher paper and code the first 7 DSM question as YES if the respondent responded “Often” and code the last 3 DSM questions as YES if the respondent responded “Once or Twice,” “Sometimes” or “Often.” Following the DSM-IV, someone is declared to be a Pathological Gambler if they have 5 or more positive DSM responses.

Using these criteria we have the following:

```
  dsm_score |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      1,612       96.47       96.47
          1 |         44        2.63       99.10
          2 |          6        0.36       99.46
          3 |          5        0.30       99.76
          4 |          2        0.12       99.88
          5 |          2        0.12      100.00
------------+-----------------------------------
      Total |      1,671      100.00
```

D. GACS

We have 1,626 valid survey respondents that completed the GACS survey questions, and of course everyone in that 1,626 also did FLAGS.

Here is the Likert scale used for each question, with one (GACS_5) reverse-coded:

```
   If it were possible, I |
    probably would gamble |
            now.          |      Freq.     Percent        Cum.
--------------------------+-----------------------------------
        strongly disagree |      1,152       70.85       70.85
      moderately disagree |        145        8.92       79.77
          mildly disagree |         70        4.31       84.07
neither agree or disagree |        127        7.81       91.88
             mildly agree |         75        4.61       96.49
         moderately agree |         31        1.91       98.40
           strongly agree |         26        1.60      100.00
--------------------------+-----------------------------------
                    Total |      1,626      100.00
```

```
 If it were |
possible, I |
   probably |
      would |
gamble now. |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      1,152       70.85       70.85
          2 |        145        8.92       79.77
          3 |         70        4.31       84.07
          4 |        127        7.81       91.88
          5 |         75        4.61       96.49
          6 |         31        1.91       98.40
          7 |         26        1.60      100.00
------------+-----------------------------------
      Total |      1,626      100.00
```

So the usual way one generates a score from such things is just to average. They do discuss 3 subscales from the 18 questions, but we do not want to presume factor loadings from their statistical analysis would apply to our analysis or population.

Here is the raw tabulation of the scale, rounded to the nearest 0.1 and 1 for ease of viewing:

```
gacs_scoreR |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        526       32.35       32.35
        1.1 |        107        6.58       38.93
        1.2 |        133        8.18       47.11
        1.3 |        232       14.27       61.38
        1.4 |         65        4.00       65.38
        1.5 |         60        3.69       69.07
        1.6 |         53        3.26       72.32
        1.7 |         53        3.26       75.58
        1.8 |         49        3.01       78.60
        1.9 |         29        1.78       80.38
          2 |         19        1.17       81.55
        2.1 |         31        1.91       83.46
        2.2 |         35        2.15       85.61
        2.3 |         25        1.54       87.15
        2.4 |         28        1.72       88.87
        2.5 |         11        0.68       89.54
        2.6 |         15        0.92       90.47
        2.7 |         12        0.74       91.21
        2.8 |         15        0.92       92.13
        2.9 |         10        0.62       92.74
          3 |          4        0.25       92.99
        3.1 |          6        0.37       93.36
        3.2 |         10        0.62       93.97
        3.3 |          8        0.49       94.46
        3.4 |          8        0.49       94.96
        3.5 |          5        0.31       95.26
        3.6 |          4        0.25       95.51
        3.7 |          5        0.31       95.82
        3.8 |          3        0.18       96.00
        3.9 |          6        0.37       96.37
          4 |         27        1.66       98.03
        4.1 |          6        0.37       98.40
        4.2 |          2        0.12       98.52
        4.3 |          2        0.12       98.65
        4.4 |          5        0.31       98.95
        4.5 |          2        0.12       99.08
        4.8 |          1        0.06       99.14
        5.1 |          1        0.06       99.20
        5.3 |          1        0.06       99.26
        5.4 |          1        0.06       99.32
        5.5 |          2        0.12       99.45
        5.6 |          1        0.06       99.51
        5.8 |          1        0.06       99.57
        6.1 |          1        0.06       99.63
        6.2 |          1        0.06       99.69
        6.3 |          1        0.06       99.75
        6.7 |          4        0.25      100.00
------------+-----------------------------------
      Total |      1,626      100.00
```

```
        RR |         0          1 |     Total
-----------+----------------------+----------
         1 |       476        587 |     1,063
         2 |       179        203 |       382
         3 |        40         59 |        99
         4 |        30         35 |        65
         5 |         4          2 |         6
         6 |         5          2 |         7
         7 |         2          2 |         4
-----------+----------------------+----------
     Total |       736        890 |     1,626
```

```
          Pearson chi2(6) =   4.5373   Pr = 0.604
 likelihood-ratio chi2(6) =   4.5755   Pr = 0.599
                    gamma =  -0.0236  ASE = 0.047
          Kendall's tau-b =  -0.0119  ASE = 0.024
```

E. GRCS

We have 1,656 valid survey respondents that completed the GRCS survey questions, and of course everyone in that 1,656 also did FLAGS.

Here is the Likert scale used for each question:

```
Gambling makes me happier. |      Freq.     Percent        Cum.
 -------------------------+-----------------------------------
        strongly disagree |        797       48.13       48.13
      moderately disagree |        162        9.78       57.91
          mildly disagree |         65        3.93       61.84
 neither agree or disagree |        364       21.98       83.82
             mildly agree |        190       11.47       95.29
         moderately agree |         55        3.32       98.61
           strongly agree |         23        1.39      100.00
 -------------------------+-----------------------------------
                    Total |      1,656      100.00

   Gambling |
   makes me |
   happier. |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        797       48.13       48.13
          2 |        162        9.78       57.91
          3 |         65        3.93       61.84
          4 |        364       21.98       83.82
          5 |        190       11.47       95.29
          6 |         55        3.32       98.61
          7 |         23        1.39      100.00
------------+-----------------------------------
      Total |      1,656      100.00
```

Raylu & Oei [2004b] document clearly how to score this instrument. We generate the simple average subscale score and then the simple average of those for an overall score.

Here is the raw tabulation of the scale, rounded to the nearest 0.1 and 1 for ease of viewing:

```
grcs_scoreR |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        482       29.11       29.11
        1.1 |        195       11.78       40.88
        1.2 |         98        5.92       46.80
        1.3 |        132        7.97       54.77
        1.4 |         86        5.19       59.96
        1.5 |         86        5.19       65.16
        1.6 |         80        4.83       69.99
        1.7 |         44        2.66       72.64
        1.8 |         51        3.08       75.72
        1.9 |         30        1.81       77.54
          2 |         30        1.81       79.35
        2.1 |         32        1.93       81.28
        2.2 |         36        2.17       83.45
        2.3 |         32        1.93       85.39
        2.4 |         20        1.21       86.59
        2.5 |         28        1.69       88.29
        2.6 |         14        0.85       89.13
        2.7 |         16        0.97       90.10
        2.8 |         16        0.97       91.06
        2.9 |          9        0.54       91.61
          3 |          9        0.54       92.15
        3.1 |          8        0.48       92.63
        3.2 |         16        0.97       93.60
        3.3 |          5        0.30       93.90
        3.4 |         15        0.91       94.81
        3.5 |          7        0.42       95.23
        3.6 |          3        0.18       95.41
        3.7 |         10        0.60       96.01
        3.8 |          5        0.30       96.32
        3.9 |          4        0.24       96.56
          4 |         35        2.11       98.67
        4.1 |          5        0.30       98.97
        4.2 |          2        0.12       99.09
        4.3 |          2        0.12       99.21
        4.6 |          1        0.06       99.28
        4.7 |          3        0.18       99.46
        4.8 |          1        0.06       99.52
        4.9 |          1        0.06       99.58
        5.2 |          1        0.06       99.64
        5.5 |          1        0.06       99.70
          6 |          1        0.06       99.76
        6.3 |          1        0.06       99.82
        6.7 |          1        0.06       99.88
        6.9 |          2        0.12      100.00
------------+-----------------------------------
      Total |      1,656      100.00
```

```
grcs_score |  Lifetime prevalence
        RR |         0          1 |     Total
-----------+----------------------+----------
         1 |       472        562 |     1,034 
         2 |       179        239 |       418 
         3 |        56         64 |       120 
         4 |        34         37 |        71 
         5 |         6          1 |         7 
         6 |         3          0 |         3 
         7 |         1          2 |         3 
-----------+----------------------+----------
     Total |       751        905 |     1,656 

          Pearson chi2(6) =   9.7742   Pr = 0.134
 likelihood-ratio chi2(6) =  11.2655   Pr = 0.081
                    gamma =  -0.0028  ASE = 0.046
          Kendall's tau-b =  -0.0014  ASE = 0.024
```

F. GUS

We have 1,695 valid survey respondents that completed the GUS survey questions, and of course everyone in that 1,695 also did FLAGS.

Here is the Likert scale used for each question:

```
  All I want to do now is |
               to gamble. |      Freq.     Percent        Cum.
--------------------------+-----------------------------------
        strongly disagree |      1,432       84.48       84.48
      moderately disagree |         89        5.25       89.73
          mildly disagree |         31        1.83       91.56
neither agree or disagree |         86        5.07       96.64
             mildly agree |         27        1.59       98.23
         moderately agree |         16        0.94       99.17
           strongly agree |         14        0.83      100.00
--------------------------+-----------------------------------
                    Total |      1,695      100.00
```

```
 All I want |
 to do now |
     is to |
   gamble. |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      1,432       84.48       84.48
          2 |         89        5.25       89.73
          3 |         31        1.83       91.56
          4 |         86        5.07       96.64
          5 |         27        1.59       98.23
          6 |         16        0.94       99.17
          7 |         14        0.83      100.00
------------+-----------------------------------
      Total |      1,695      100.00
```

Raylu & Oei [2004a] document clearly how to score this: they suggest the total score, since this is viewed as one factor, but the average score would do the same. They drop 2 of the 8 questions (the reverse-coded ones, GUS_7 and GUS_8).

Here is the raw tabulation of the scale, rounded to the nearest 0.1 and 1 for ease of viewing:

```
 gus_scoreR |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        996       58.76       58.76
        1.1 |         70        4.13       62.89
        1.3 |         50        2.95       65.84
        1.4 |         66        3.89       69.73
        1.5 |         42        2.48       72.21
        1.6 |         40        2.36       74.57
        1.8 |        126        7.43       82.01
        1.9 |         30        1.77       83.78
          2 |         23        1.36       85.13
        2.1 |         16        0.94       86.08
        2.3 |         18        1.06       87.14
        2.4 |         12        0.71       87.85
        2.5 |         63        3.72       91.56
        2.6 |         15        0.88       92.45
        2.8 |          6        0.35       92.80
        2.9 |          7        0.41       93.22
          3 |          6        0.35       93.57
        3.1 |         10        0.59       94.16
        3.3 |         10        0.59       94.75
        3.4 |          5        0.29       95.04
        3.5 |          3        0.18       95.22
        3.6 |          8        0.47       95.69
        3.8 |          9        0.53       96.22
        3.9 |          6        0.35       96.58
          4 |         23        1.36       97.94
        4.1 |          3        0.18       98.11
        4.3 |          3        0.18       98.29
        4.5 |          1        0.06       98.35
        4.6 |          4        0.24       98.58
        4.8 |          3        0.18       98.76
        4.9 |          3        0.18       98.94
          5 |          2        0.12       99.06
        5.1 |          1        0.06       99.12
        5.3 |          2        0.12       99.23
        5.4 |          1        0.06       99.29
        5.5 |          2        0.12       99.41
        5.6 |          1        0.06       99.47
        5.9 |          1        0.06       99.53
          6 |          1        0.06       99.59
        6.3 |          2        0.12       99.71
        6.5 |          1        0.06       99.76
        6.8 |          1        0.06       99.82
          7 |          3        0.18      100.00
------------+-----------------------------------
      Total |      1,695      100.00

gus_scoreRR |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      1,182       69.73       69.73
          2 |        307       18.11       87.85
          3 |        122        7.20       95.04
          4 |         55        3.24       98.29
          5 |         17        1.00       99.29
          6 |          7        0.41       99.71
          7 |          5        0.29      100.00
------------+-----------------------------------
      Total |      1,695      100.00


gus_scoreR |  Lifetime prevalence
         R |         0          1 |     Total
-----------+----------------------+----------
         1 |       505        677 |     1,182
         2 |       136        171 |       307
         3 |        59         63 |       122
         4 |        30         25 |        55
         5 |         9          8 |        17
         6 |         5          2 |         7
         7 |         5          0 |         5
-----------+----------------------+----------
     Total |       749        946 |     1,695

          Pearson chi2(6) =  13.2318   Pr = 0.039
 likelihood-ratio chi2(6) =  15.0737   Pr = 0.020
                    gamma =  -0.1055  ASE = 0.047
          Kendall's tau-b =  -0.0514  ASE = 0.023
```

G. AUDIT

We have 6,661 valid survey respondents that completed the AUDIT survey questions, and of course everyone in that 6,661 also did FLAGS.

Here is the Likert scale used for each of the first 8 questions:

```
How often during the |
   last year have you |
  found that you were |
 not able to stop dri |      Freq.     Percent        Cum.
----------------------+-----------------------------------
                Never |      5,707       85.85       85.85
    Less than monthly |        695       10.45       96.30
              Monthly |        154        2.32       98.62
               Weekly |         65        0.98       99.59
Daily or almost daily |         27        0.41      100.00
----------------------+-----------------------------------
                Total |      6,648      100.00
```

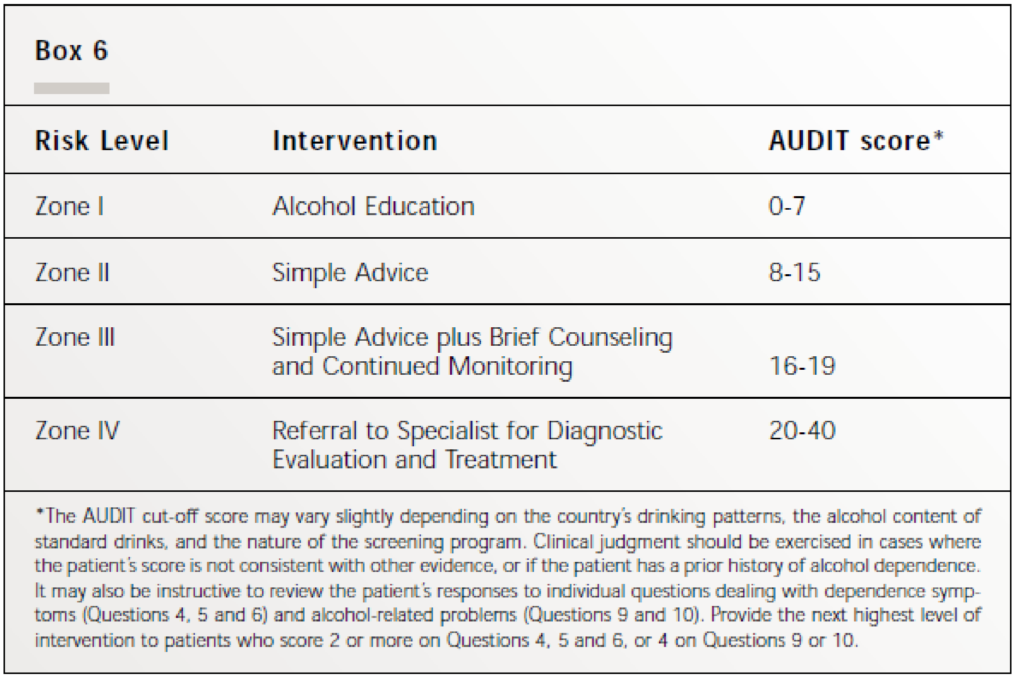
```
  How often |
 during the |
  last year |
   have you |
 found that |
   you were |
not able to |
   stop dri |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      5,707       85.85       85.85
          1 |        695       10.45       96.30
          2 |        154        2.32       98.62
          3 |         65        0.98       99.59
          4 |         27        0.41      100.00
------------+-----------------------------------
      Total |      6,648      100.00
```

But the last 2 questions have different coding, with the numerical values shown:

```
Have you or someone else been |
      injured because of your |
                 drinking?    |      Freq.     Percent        Cum.
------------------------------+-----------------------------------
                           No |      6,071       91.32       91.32
Yes, but not in the last year |        467        7.02       98.35
    Yes, during the last year |        110        1.65      100.00
------------------------------+-----------------------------------
                        Total |      6,648      100.00
```

```
Have you or |
    someone |
  else been |
    injured |
 because of |
       your |
drinking?   |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      6,071       91.32       91.32
          2 |        467        7.02       98.35
          4 |        110        1.65      100.00
------------+-----------------------------------
      Total |      6,648      100.00
```

Bablor et al. [2001] document clearly how to score this: they suggest the total score, since this is viewed as one factor. While the average score would do the same, we want comparability to cutoff thresholds stated in terms of the score. For instance, here is the default categories based on the total score:



Box 6

Risk Level	**Intervention**	**AUDIT score***
Zone I	Alcohol Education	0-7
Zone II	Simple Advice	8-15
Zone III	Simple Advice plus Brief Counseling and Continued Monitoring	16-19
Zone IV	Referral to Specialist for Diagnostic Evaluation and Treatment	20-40

*The AUDIT cut-off score may vary slightly depending on the country's drinking patterns, the alcohol content of standard drinks, and the nature of the screening program. Clinical judgment should be exercised in cases where the patient's score is not consistent with other evidence, or if the patient has a prior history of alcohol dependence. It may also be instructive to review the patient's responses to individual questions dealing with dependence symptoms (Questions 4, 5 and 6) and alcohol-related problems (Questions 9 and 10). Provide the next highest level of intervention to patients who score 2 or more on Questions 4, 5 and 6, or 4 on Questions 9 or 10.

Here is the raw tabulation of the scale, with extra lines between the above "zones":

```
audit_score |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        406        6.11        6.11
          1 |        575        8.65       14.76
          2 |        712       10.71       25.47
          3 |        928       13.96       39.43
          4 |        998       15.01       54.44
          5 |        711       10.69       65.13
          6 |        556        8.36       73.50
          7 |        386        5.81       79.30
------------+-----------------------------------
          8 |        313        4.71       84.01
          9 |        212        3.19       87.20
         10 |        169        2.54       89.74
         11 |        148        2.23       91.97
         12 |         87        1.31       93.28
         13 |        103        1.55       94.83
         14 |         66        0.99       95.82
         15 |         45        0.68       96.50
------------+-----------------------------------
         16 |         50        0.75       97.25
         17 |         36        0.54       97.79
         18 |         26        0.39       98.18
         19 |         19        0.29       98.47
------------+-----------------------------------
         20 |         21        0.32       98.78
         21 |         14        0.21       98.99
         22 |         19        0.29       99.28
         23 |          8        0.12       99.40
         24 |          7        0.11       99.50
         25 |          9        0.14       99.64
         26 |          3        0.05       99.68
         27 |          4        0.06       99.74
         28 |          2        0.03       99.77
         29 |          3        0.05       99.82
         30 |          1        0.02       99.83
         31 |          1        0.02       99.85
         32 |          3        0.05       99.89
         33 |          2        0.03       99.92
         34 |          3        0.05       99.97
         35 |          1        0.02       99.98
         40 |          1        0.02      100.00
------------+-----------------------------------
      Total |      6,648      100.00
```

Here are the evaluations of treatments:

```
audit_scor |  Lifetime prevalence
         e |         0          1 |     Total
-----------+----------------------+----------
         0 |       188        218 |       406 
         1 |       263        312 |       575 
         2 |       306        406 |       712 
         3 |       393        535 |       928 
         4 |       431        567 |       998 
         5 |       326        385 |       711 
         6 |       241        315 |       556 
         7 |       177        209 |       386 
         8 |       142        171 |       313 
         9 |       101        111 |       212 
        10 |        77         92 |       169 
        11 |        62         86 |       148 
        12 |        49         38 |        87 
        13 |        50         53 |       103 
        14 |        27         39 |        66 
        15 |        26         19 |        45 
        16 |        24         26 |        50 
        17 |        21         15 |        36 
        18 |        13         13 |        26 
        19 |         9         10 |        19 
        20 |        13          8 |        21 
        21 |        11          3 |        14 
        22 |        13          6 |        19 
        23 |         5          3 |         8 
        24 |         3          4 |         7 
        25 |         4          5 |         9 
        26 |         1          2 |         3 
        27 |         1          3 |         4 
        28 |         0          2 |         2 
        29 |         1          2 |         3 
        30 |         0          1 |         1 
        31 |         0          1 |         1 
        32 |         2          1 |         3 
        33 |         0          2 |         2 
        34 |         2          1 |         3 
        35 |         0          1 |         1 
        40 |         0          1 |         1 
-----------+----------------------+----------
     Total |     2,982      3,666 |     6,648 

          Pearson chi2(36) =  41.8557   Pr = 0.232
 likelihood-ratio chi2(36) =  45.1015   Pr = 0.142
                     gamma =  -0.0288  ASE = 0.016
           Kendall's tau-b =  -0.0193  ASE = 0.010
```

H. BAI

We have 1,671 valid survey respondents that completed the BAI survey questions, and of course everyone in that 1,671 also did FLAGS.

Here is the Likert scale used for each question:

```
                      Fear of dying   |      Freq.     Percent        Cum.
--------------------------------------+-----------------------------------
                          Not at all |      1,253       74.99       74.99
   Mildly but it didn't bother me much. |        267       15.98       90.96
Moderately - it wasn't pleasant at time |        112        6.70       97.67
       Severely - it bothered me a lot. |         39        2.33      100.00
--------------------------------------+-----------------------------------
                               Total |      1,671      100.00
```

```
   Fear of |
   dying   |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      1,253       74.99       74.99
          2 |        267       15.98       90.96
          3 |        112        6.70       97.67
          4 |         39        2.33      100.00
------------+-----------------------------------
      Total |      1,671      100.00
```

Beck and Steer [1990] document clearly how to score this: they suggest the total score, since this is viewed as one factor, but the average score would do the same. We have 21 BAI questions.   Here is the raw tabulation of the scale, rounded to the nearest 0.1 for ease of viewing:

```
 bai_scoreR |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        537       32.14       32.14
        1.1 |        159        9.52       41.65
        1.2 |        139        8.32       49.97
        1.3 |        130        7.78       57.75
        1.4 |        105        6.28       64.03
        1.5 |         90        5.39       69.42
        1.6 |         88        5.27       74.69
        1.7 |         58        3.47       78.16
        1.8 |         50        2.99       81.15
        1.9 |         47        2.81       83.96
          2 |         51        3.05       87.01
        2.1 |         41        2.45       89.47
        2.2 |         40        2.39       91.86
        2.3 |         16        0.96       92.82
        2.4 |         22        1.32       94.14
        2.5 |         12        0.72       94.85
        2.6 |         14        0.84       95.69
        2.7 |          9        0.54       96.23
        2.8 |         10        0.60       96.83
        2.9 |         10        0.60       97.43
          3 |         14        0.84       98.26
        3.1 |          8        0.48       98.74
        3.2 |          2        0.12       98.86
        3.3 |          1        0.06       98.92
        3.4 |          7        0.42       99.34
        3.5 |          4        0.24       99.58
        3.6 |          2        0.12       99.70
        3.7 |          1        0.06       99.76
        3.8 |          1        0.06       99.82
          4 |          3        0.18      100.00
```

```
------------+-----------------------------------
      Total |      1,671      100.00
```

Here are the evaluations of treatments:

```
           |  Lifetime prevalence
bai_scoreR |         0          1 |     Total
-----------+----------------------+----------
         1 |       245        292 |       537 
       1.1 |        72         87 |       159 
       1.2 |        60         79 |       139 
       1.3 |        53         77 |       130 
       1.4 |        58         47 |       105 
       1.5 |        33         57 |        90 
       1.6 |        45         43 |        88 
       1.7 |        28         30 |        58 
       1.8 |        18         32 |        50 
       1.9 |        22         25 |        47 
         2 |        22         29 |        51 
       2.1 |        17         24 |        41 
       2.2 |        16         24 |        40 
       2.3 |         6         10 |        16 
       2.4 |        11         11 |        22 
       2.5 |         5          7 |        12 
       2.6 |         5          9 |        14 
       2.7 |         4          5 |         9 
       2.8 |         3          7 |        10 
       2.9 |         5          5 |        10 
         3 |         7          7 |        14 
       3.1 |         5          3 |         8 
       3.2 |         1          1 |         2 
       3.3 |         0          1 |         1 
       3.4 |         2          5 |         7 
       3.5 |         2          2 |         4 
       3.6 |         0          2 |         2 
       3.7 |         0          1 |         1 
       3.8 |         0          1 |         1 
         4 |         1          2 |         3 
-----------+----------------------+----------
     Total |       746        925 |     1,671 

          Pearson chi2(29) =  20.4571   Pr = 0.878
 likelihood-ratio chi2(29) =  22.4397   Pr = 0.802
                     gamma =   0.0250  ASE = 0.032
           Kendall's tau-b =   0.0163  ASE = 0.021
```

I. BDI

We have 1,695 valid survey respondents that completed the BDI survey questions, and of course everyone in that 1,695 also did FLAGS.

Here is the Likert scale used for each question:

```
               Loss of Interest in Sex |      Freq.     Percent        Cum.
---------------------------------------+-----------------------------------
       I have not noticed any recent ch |      1,159       68.38       68.38
       I am less interested in sex than |        316       18.64       87.02
       I am much less interested in sex |        117        6.90       93.92
       I have lost interest in sex comp |        103        6.08      100.00
---------------------------------------+-----------------------------------
```

```
                                Total |      1,695      100.00

    Loss of |
Interest in |
        Sex |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      1,159       68.38       68.38
          1 |        316       18.64       87.02
          2 |        117        6.90       93.92
          3 |        103        6.08      100.00
------------+-----------------------------------
      Total |      1,695      100.00
```

Beck et al. [1961] document clearly how to score this: they suggest the total score, since this is viewed as one factor, but the average score would do the same.

Here is the raw tabulation of the scale, rounded to the nearest 0.1 for ease of viewing:

```
 bdi_scoreR |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        631       37.23       37.23
         .1 |        167        9.85       47.08
         .2 |        110        6.49       53.57
         .3 |         67        3.95       57.52
         .4 |         52        3.07       60.59
         .5 |         35        2.06       62.65
         .6 |         44        2.60       65.25
         .7 |         63        3.72       68.97
         .8 |         51        3.01       71.98
         .9 |         50        2.95       74.93
          1 |         50        2.95       77.88
        1.1 |         34        2.01       79.88
        1.2 |         27        1.59       81.47
        1.3 |         33        1.95       83.42
        1.4 |         26        1.53       84.96
        1.5 |         18        1.06       86.02
        1.6 |         19        1.12       87.14
        1.7 |         23        1.36       88.50
        1.8 |         25        1.47       89.97
        1.9 |         15        0.88       90.86
          2 |         16        0.94       91.80
        2.1 |         13        0.77       92.57
        2.2 |         16        0.94       93.51
        2.3 |         10        0.59       94.10
        2.4 |         13        0.77       94.87
        2.5 |          8        0.47       95.34
        2.6 |          7        0.41       95.75
        2.7 |         10        0.59       96.34
        2.8 |          7        0.41       96.76
        2.9 |          7        0.41       97.17
          3 |          9        0.53       97.70
        3.1 |          6        0.35       98.05
        3.2 |          6        0.35       98.41
        3.3 |          4        0.24       98.64
        3.4 |          5        0.29       98.94
        3.5 |          3        0.18       99.12
        3.6 |          3        0.18       99.29
        3.7 |          1        0.06       99.35
        3.8 |          4        0.24       99.59
        3.9 |          2        0.12       99.71
          4 |          4        0.24       99.94
        4.2 |          1        0.06      100.00
------------+-----------------------------------
      Total |      1,695      100.00
```

J. BIS

We have 1,761 valid survey respondents that completed the BIS survey questions, and of course everyone in that 1,761 also did FLAGS.

Here is the Likert scale used for each question:

```
          I plan tasks |
   carefully.          |      Freq.     Percent        Cum.
-----------------------+-----------------------------------
                 Never |         74        4.21        4.21
          Occasionally |        494       28.12       32.33
                 Often |        786       44.74       77.06
Almost always / Always |        403       22.94      100.00
-----------------------+-----------------------------------
                 Total |      1,757      100.00
```

```
     I plan |
      tasks |
 carefully. |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |         74        4.21        4.21
          2 |        494       28.12       32.33
          3 |        786       44.74       77.06
          4 |        403       22.94      100.00
------------+-----------------------------------
      Total |      1,757      100.00
```

Patton, Stanford and Barratt [1995] document clearly how to score this: they suggest the total score, since this is viewed as one factor, but the average score would do the same. We have 30 BIS questions.

Here is the raw tabulation of the scale, rounded to the nearest 0.1 for ease of viewing:

```
bis_scoreR |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |          3        0.17        0.17
        1.1 |          2        0.11        0.28
        1.2 |          3        0.17        0.46
        1.3 |          1        0.06        0.51
        1.4 |          3        0.17        0.68
        1.5 |          3        0.17        0.85
        1.6 |          6        0.34        1.20
        1.7 |         52        2.96        4.15
        1.8 |        116        6.60       10.76
        1.9 |        214       12.18       22.94
          2 |        291       16.56       39.50
        2.1 |        366       20.83       60.33
        2.2 |        309       17.59       77.92
        2.3 |        209       11.90       89.81
        2.4 |        102        5.81       95.62
        2.5 |         53        3.02       98.63
        2.6 |         17        0.97       99.60
        2.7 |          3        0.17       99.77
        2.8 |          3        0.17       99.94
        2.9 |          1        0.06      100.00
------------+-----------------------------------
      Total |      1,757      100.00
```

Appendix C: Summary Statistics

Of the 65,592 Danes that were invited to participate in the survey, only 8,405 completed the entire survey, while 3,331 started but did not complete the survey. We here present the background information that is available for participants and non-participants.

We have background information for all participants and non-participants on sex, age, and region, but we have missing values for the other variables for several non-participants. We received background information on non-participants from the companies that administered the survey, and they could not provide complete information for all invited people.

	Completed					**Partially Completed**					**Non-Participants**				
	N	*Mean*	*Std. Dev.*	*Min*	*Max*	*N*	*Mean*	*Std. Dev.*	*Min*	*Max*	*N*	*Mean*	*Std. Dev.*	*Min*	*Max*
Female	8405	0.54	0.50	0	1	3331	0.58	0.49	0	1	53856	0.62	0.49	0	1
Age	8405	49.19	14.82	18	75	3331	48.15	15.23	18	75	53856	41.32	14.07	18	75
Young (18-29)	8405	0.13	0.33	0	1	3331	0.16	0.36	0	1	53856	0.24	0.42	0	1
Middle (30-39)	8405	0.15	0.36	0	1	3331	0.15	0.35	0	1	53856	0.23	0.42	0	1
Ripe Aged (40-49)	8405	0.20	0.40	0	1	3331	0.20	0.40	0	1	53856	0.25	0.43	0	1
Older (50+)	8405	0.52	0.50	0	1	3331	0.49	0.50	0	1	53856	0.29	0.45	0	1
Capital Region	8405	0.49	0.50	0	1	3331	0.55	0.50	0	1	53856	0.44	0.50	0	1
Central Denmark	8405	0.17	0.37	0	1	3331	0.15	0.36	0	1	53856	0.18	0.39	0	1
North Denmark	8405	0.06	0.25	0	1	3331	0.06	0.23	0	1	53856	0.07	0.26	0	1
Zealand	8405	0.12	0.33	0	1	3331	0.11	0.31	0	1	53856	0.13	0.34	0	1
Southern Denmark	8405	0.16	0.37	0	1	3331	0.13	0.33	0	1	53856	0.17	0.38	0	1
Lives in a House	8405	0.56	0.50	0	1	3000	0.55	0.50	0	1	42699	0.55	0.50	0	1
Owns a House	8405	0.55	0.50	0	1	3030	0.56	0.50	0	1	43046	0.53	0.50	0	1
Lives with a Partner	8405	0.64	0.48	0	1	2991	0.64	0.48	0	1	38373	0.63	0.48	0	1
Children in Household	8405	0.26	0.44	0	1	944	0.25	0.43	0	1	9177	0.43	0.50	0	1
Civil Servant	8405	0.39	0.49	0	1	3094	0.44	0.50	0	1	44428	0.43	0.49	0	1
Retired	8405	0.23	0.42	0	1	3094	0.17	0.38	0	1	44428	0.08	0.27	0	1
Unskilled	8405	0.05	0.22	0	1	3094	0.04	0.20	0	1	44428	0.06	0.24	0	1
Skilled	8405	0.10	0.30	0	1	3094	0.08	0.28	0	1	44428	0.12	0.33	0	1
Self Employed	8405	0.05	0.21	0	1	3094	0.04	0.20	0	1	44428	0.04	0.20	0	1
Vocational Training	8405	0.23	0.42	0	1	2993	0.19	0.39	0	1	37716	0.19	0.39	0	1
Low Formal Education	8405	0.21	0.41	0	1	2993	0.21	0.41	0	1	37716	0.26	0.44	0	1
Short Education (<3 Y. College)	8405	0.10	0.31	0	1	2993	0.09	0.29	0	1	37716	0.09	0.28	0	1
Med. Education (3-4 Y. College)	8405	0.29	0.45	0	1	2993	0.31	0.46	0	1	37716	0.30	0.46	0	1
High Education (5+ Y. College)	8405	0.17	0.38	0	1	2993	0.18	0.38	0	1	37716	0.15	0.36	0	1
Low Income (<300,000)	8405	0.37	0.48	0	1	3033	0.36	0.48	0	1	43021	0.42	0.49	0	1
Med. Income (300,000-500,000)	8405	0.34	0.47	0	1	3033	0.33	0.47	0	1	43021	0.32	0.47	0	1
High Income (500,000-800,000)	8405	0.12	0.33	0	1	3033	0.12	0.32	0	1	43021	0.10	0.30	0	1
Very High Income (>800,000)	8405	0.02	0.15	0	1	3033	0.03	0.16	0	1	43021	0.02	0.15	0	1
Silent about Income	8405	0.14	0.35	0	1	3033	0.16	0.37	0	1	43021	0.14	0.34	0	1
Q: Experienced Death	8405	0.12	0.33	0	1	33	0.12	0.33	0	1	0				
Q: Has been Hospitalized	8405	0.10	0.30	0	1	33	0.06	0.24	0	1	0				
Q: Trigger 40 DKK	8405	0.38	0.49	0	1	33	0.24	0.44	0	1	0				
Q: Trigger 500 DKK	8405	0.05	0.22	0	1	33	0.03	0.17	0	1	0				
Q: Currently Smoker	8405	0.20	0.40	0	1	32	0.34	0.48	0	1	0				

Note: The last 5 variables are based on additional questions asked in the surveys (see Appendix A).

Appendix D: Additional Figures on Marginal Effects of Demographic Characteristics

We present additional figures that illustrate the marginal effects of demographic characteristics on FLAGS levels. The estimations are based on a semi-nonparametric ordered response model with controls for sample selection and sample weights.

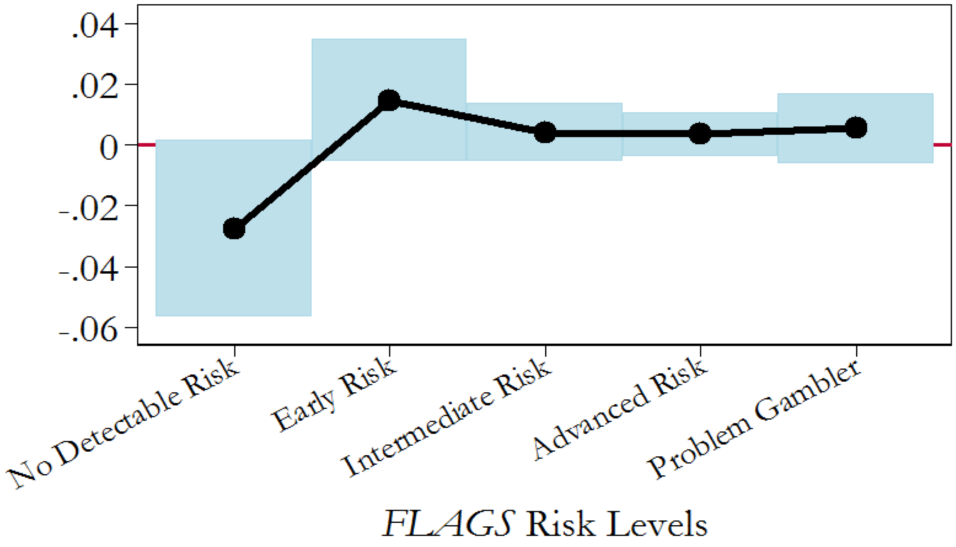


Figure D1: Marginal Effect of Being Young
on Probability of *FLAGS* Gambling Risk Level

Young defined as aged between 18 and 29
Semi-Nonparametric Ordered Response model with
sample weights and sample selection corrections
Point estimate of effect and 95% confidence interval

Figure D2: Marginal Effect of Being Ripe_Aged on Probability of *FLAGS* Gambling Risk Level

Ripe-Aged defined as aged between 40 and 49
Semi-Nonparametric Ordered Response model with
sample weights and sample selection corrections
Point estimate of effect and 95% confidence interval

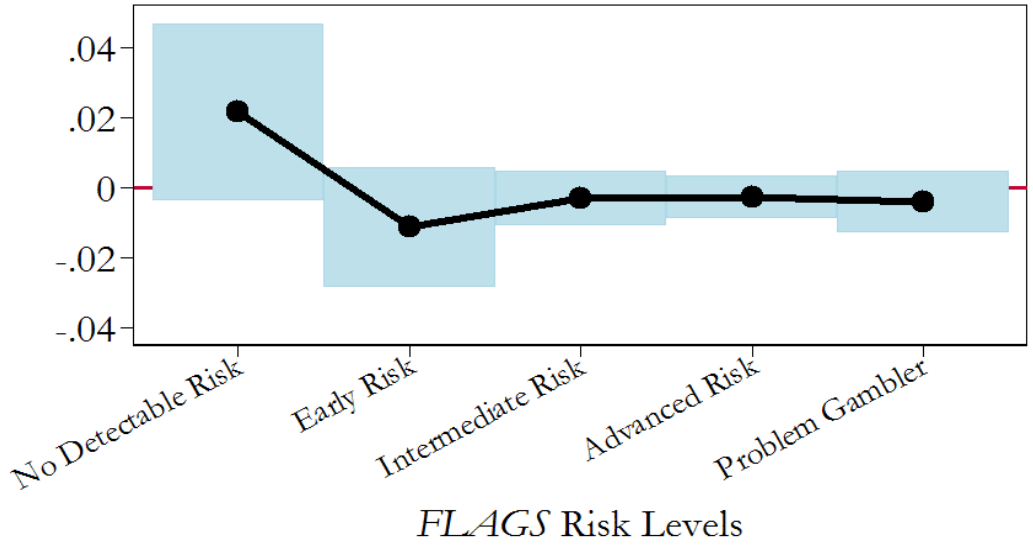
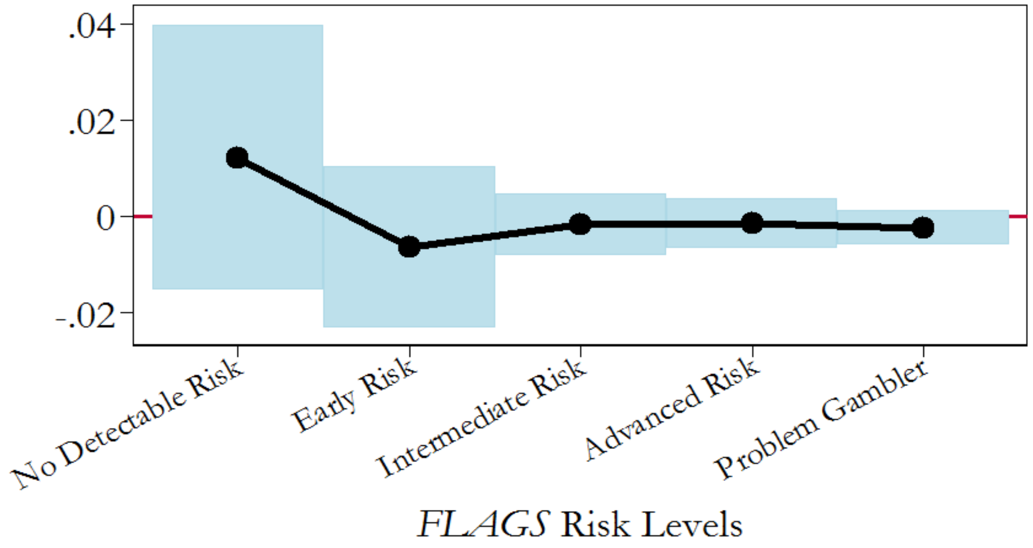


Figure D3: Marginal Effect of Being Older on Probability of *FLAGS* Gambling Risk Level

Older defined as aged 50 and over
Semi-Nonparametric Ordered Response model with
sample weights and sample selection corrections
Point estimate of effect and 95% confidence interval



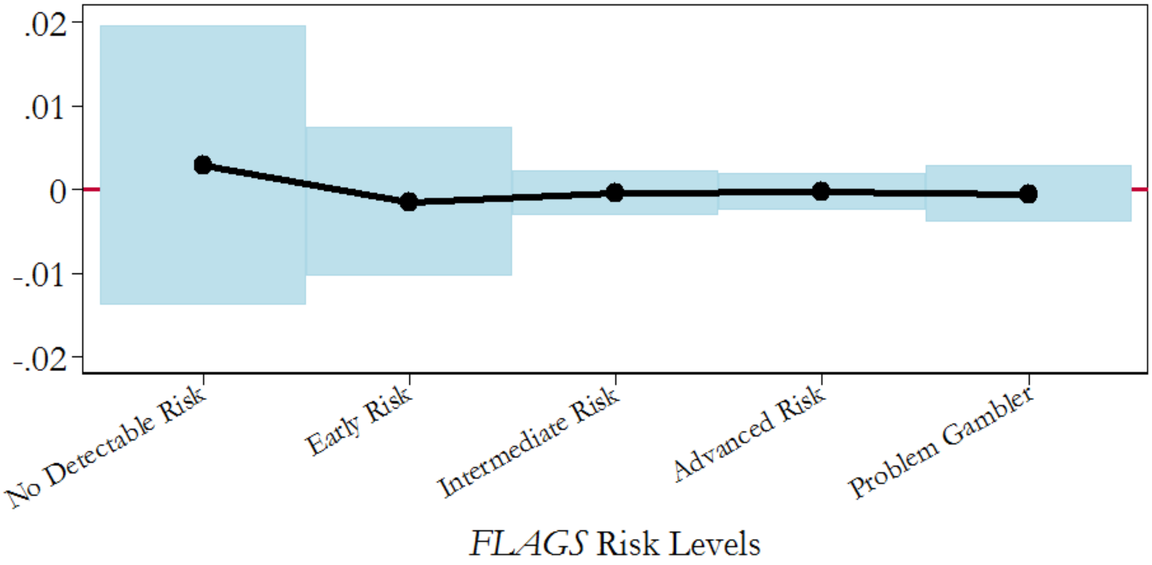


Figure D4: Marginal Effect of Having Medium Income on Probability of *FLAGS* Gambling Risk Level

Medium Income defined by income between 300,000 and 500,000 kroner p.a.
Semi-Nonparametric Ordered Response model with
sample weights and sample selection corrections
Point estimate of effect and 95% confidence interval

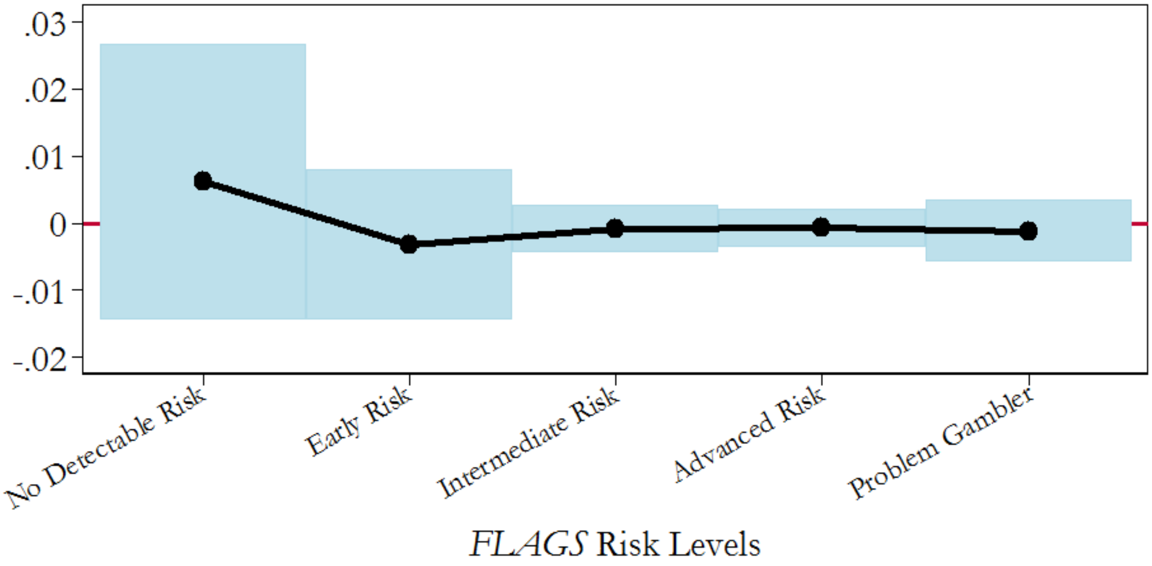


Figure D5: Marginal Effect of Having High Income on Probability of *FLAGS* Gambling Risk Level

Medium Income defined by income between 500,000 and 800,000 kroner p.a.
Semi-Nonparametric Ordered Response model with
sample weights and sample selection corrections
Point estimate of effect and 95% confidence interval

Figure D6: Marginal Effect of Having Very High Income
on Probability of *FLAGS* Gambling Risk Level

Very High Income defined by income between over 800,000 kroner p.a.
Semi-Nonparametric Ordered Response model with
sample weights and sample selection corrections
Point estimate of effect and 95% confidence interval

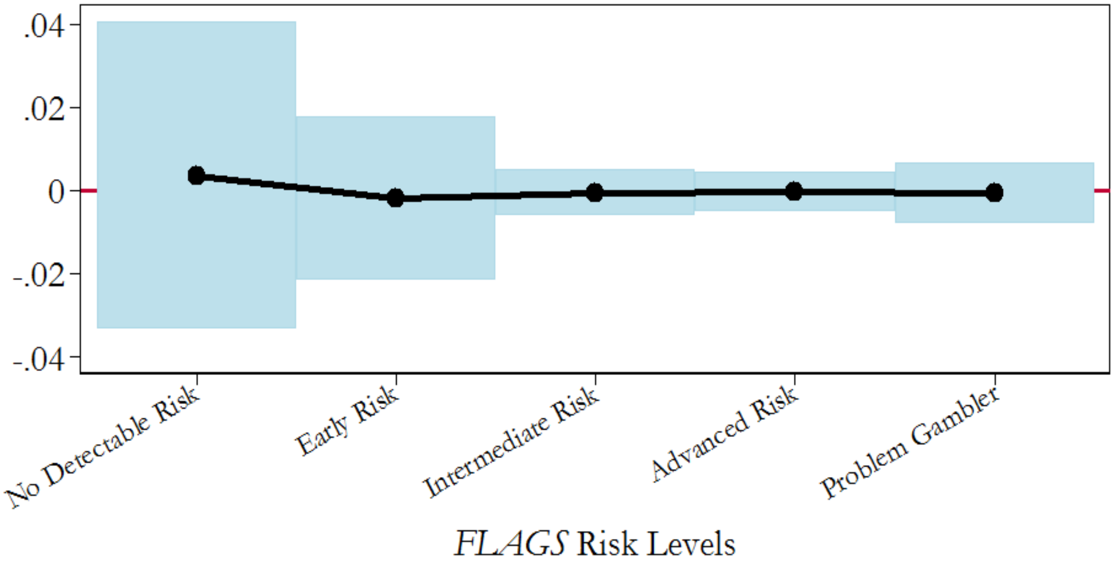
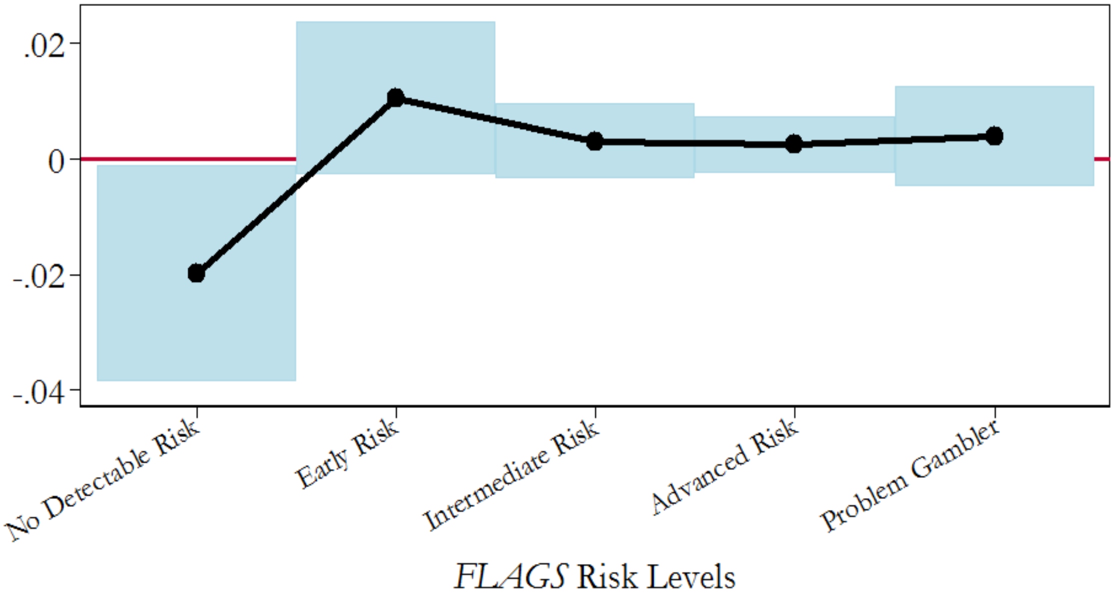


Figure D7: Marginal Effect of Being a Current Smoker
on Probability of *FLAGS* Gambling Risk Level

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