



Published in final edited form as:

Assessment. 2013 October ; 20(5): 523–531. doi:10.1177/1073191113500522.

Evaluating the South Oaks Gambling Screen With *DSM-IV* and *DSM-5* Criteria: Results From a Diverse Community Sample of Gamblers

Adam S. Goodie¹, James MacKillop^{1,2}, Joshua D. Miller¹, Erica E. Fortune¹, Jessica Maples¹, Charles E. Lance¹, and W. Keith Campbell¹

¹University of Georgia, Athens, GA, USA

²Brown University, Providence, RI, USA

Abstract

Despite widespread use, the South Oaks Gambling Screen (SOGS) has been criticized for excessive false positives as an indicator of pathological gambling (PG), and for items that misalign with PG criteria. We examine the relationship between SOGS scores and PG symptoms and convergent validity with regard to personality, mood, and addictive behaviors in a sample of 353 gamblers. SOGS scores correlated $r = .66$ with both *DSM-IV* and *DSM-5* symptoms, and they manifested similar correlations with external criteria (intraclass correlation of .95). However, 195 false positives and 1 false negative were observed when using the recommended cut point, yielding an 81% false alarm rate. For uses with *DSM-IV* criteria, a cut point of 10 would retain high sensitivity with greater specificity and fewer false positives. For *DSM-5* criteria, we advocate a cut point of 8 for use as a clinical screen and a cut point of 12 for prevalence and pseudo-experimental studies.

Keywords

gambling; pathological gambling; South Oaks Gambling Screen; DSM-5; false alarm rate

The South Oaks Gambling Screen (SOGS; Lesieur & Blume, 1987) is a 20-item multiple-choice instrument that was introduced as a method for identifying individuals with pathological gambling (PG). Positive responses to 5 or more items result in a designation of “probable pathological gambler” (PPG; Lesieur & Blume, 1987). The SOGS’s format permits many modes of administration, including interviews conducted by either experts or nonexperts, computer, or self-administration. Because of this convenience and efficiency, the SOGS quickly became the dominant instrument for measuring PG in research settings, including applications to theoretical, clinical, and epidemiological research (Petry, 2005).

Reprints and permissions: sagepub.com/journalsPermissions.nav

Corresponding Author: Adam S. Goodie, Department of Psychology, University of Georgia, 125 Baldwin Street, Athens, GA 30602-3013, USA. goodie@uga.edu.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

The funding agencies played no role in the study design, data collection, data analysis, or preparation of the article.

Despite its widespread use, the SOGS has been subject to criticism. One objection is that the SOGS has revealed prevalence rates of PPG that are higher than diagnostic rates based on *DSM* criteria in similar populations (e.g., Stinchfield, 2002; Wickwire, Burke, Brown, Parker, & May, 2008), raising concerns about excessive false positive rates. As its name implies, the SOGS was initially developed as a screening instrument, intended to minimize false negatives, to be followed by diagnostic interviews. However, when used on its own for diagnostic purposes, assignment to pseudo-experimental groups, or prevalence studies, the false positive rate becomes a significant concern. Such uses have become quite common, beginning with prevalence results reported in the original study (Lesieur & Blume, 1987).

Other criticisms relate to the SOGS's items being poorly aligned with current *DSM* criteria (Stinchfield, 2002). Some SOGS items are subjective whereas a corresponding *DSM* criterion is behavioral, and half of the SOGS items relate to borrowing money to support gambling behavior. As Stinchfield (2002) noted, an individual could be identified as PPG with only the single symptom of relying on others to provide money for gambling. Finally, some *DSM-IV* PG symptoms are not covered by the SOGS.

These criticisms bear revisiting as the fifth edition of the *DSM* is implemented following its recent publication. *DSM-5* reduces the number of symptoms required for a PG diagnosis from 5 to 4 and eliminates 1 of the 10 criterion symptoms from *DSM-IV*, namely, the commission of illegal acts to finance gambling activity. As the diagnostic symptom-count criterion is reduced, problems of false positives are expected to diminish. It is thus timely to inquire whether the new *DSM-5* criterion will render SOGS less prone to high false positive rates and systematic overestimation of prevalence rates. In this article, we present new data that shed light on these questions, drawn from a single population of community-based frequent gamblers. We also examine the convergent validity of the SOGS based on its relations with a number of external criteria, including measures of demographic variables, personality traits, affect, and substance dependence. Finally, because the sample included robust subsamples of both White and African American subsamples, and race has been previously shown to be a significant predictor of gambling severity, we report analyses bearing on whether race is an important consideration in our conclusions.

False Positives, the False Alarm Rate, and SOGS

False positives are occurrences of positive test results (here dubbed SOGS⁺) where the underlying condition is absent (PG⁻). False positives are a component of both the *false alarm rate* (FAR) and the *false discovery rate* (FDR). The FAR, which is sometimes called the false positive rate (and is also the inverse of specificity), is defined as false positives/(false positives + true negatives), or in other words false positives as a proportion of all PG⁻ cases. Because it depends only on PG⁻ individuals, not in comparison with PG⁺ individuals, the FAR is not sensitive to base rates and therefore is not expected to vary systematically in populations with diverse PG prevalence rates. The FDR is the inverse of positive predictive value (PPV) and is defined as false positives/(false positives + true positives), or in other words false positives as a proportion of all SOGS⁺ cases. Because the FDR incorporates counts of both PG⁺ and PG⁻ individuals, it is sensitive to base rates; specifically, it is expected to be lower in populations with greater PG prevalence.

Lesieur and Blume (1987) originally reported results from a Gambler's Anonymous (GA) population and a student population, among others. Among the GA population, there were 3 true negatives, 1 false negative, 3 false positives, and 206 true positives, leading to $FAR = 3/(3 + 3) = 50\%$ and $FDR = 3/(3 + 206) = 1.4\%$. Clearly, the two uses of false positives, FAR and FDR, can have very different values for the same instrument, even within the same data set. Among the student population, which had a much lower PG rate than the GA population, there were 351 true negatives, 13 false negatives, 5 false positives, and 15 true positives, leading to $FAR = 5/(5 + 351) = 1.4\%$ and $FDR = 5/(5 + 15) = 25\%$. Lesieur and Blume (1987) and others have used multiple samples, some with large PG base rates and some with large nonpathological gambling (NPG) base rates. However, it is preferable to sample from a single population with robust subsamples in all relevant cells, often by selecting a population with moderate base rates of the target characteristic. Simultaneous sampling from multiple populations can lead to bimodal distributions that do not reflect underlying population characteristics, and in particular that exclude subclinical levels of gambling disorder.

One remedy that has been proposed to reduce SOGS's false alarm rate (e.g., Duvarci, Varan, Coskunol, & Ersoy, 1997) is an increase in the cut point. The cut point of 5 or more indicating PPG was originally justified as follows: "A score of 5 or more ... was chosen as an indication of probable pathological gambling to reduce the number of false-positive and false-negative codings" (Lesieur & Blume, 1987, p. 1186). Modifications to the cut point do not require revisions to test items, and permit reanalysis of data both collected either prior to or subsequent to changes in the cut point.

The Content of the SOGS

A second class of criticism of the SOGS is that its content does not align closely with *DSM* criteria for PG. For example, Stinchfield (2002) concluded that only 4 of the 10 *DSM-IV* criteria received any coverage in SOGS—chasing losses in Item 4, lying in Items 5 and 11, jeopardizing job or career opportunity in Item 15, and relying on others to provide money in Items 16a through 16i. Slutske, Zhu, Meier, and Martin (2011) do not count Item 15 as a match, and thus count only 3 overlapping symptoms. Furthermore, relying on others to provide gambling money is clearly overrepresented in the SOGS, as it is reflected in 10 out of the 20 items.

Several alternative scoring protocols have the potential to be useful for improving the utility of the SOGS. One is *conservative symptom matching*, which reflects Stinchfield's (2002) contention that only four symptoms are covered in the SOGS and counts positive responses to items reflecting these four symptoms. A second is *permissive symptom matching*, which uses an approach similar to conservative symptom matching but includes more symptoms. For example, Stinchfield (2002) excluded SOGS items that refer to subjective feelings rather than observable behaviors, but Lesieur and Blume (1987) argued that these subjective items may nonetheless reflect important aspects of symptomatology. This approach includes the four symptoms noted above and also preoccupation, reflected in SOGS Item 6; increasing amounts of money (tolerance), reflected in SOGS Item 7; and unsuccessful attempts to cut

down, reflected in SOGS Item 10. This increases the number of symptoms reflected from 4 to 7. Escape, restlessness when cutting down, and illegal activities remain unrepresented.

A third approach is based on an item response theory (IRT) analysis of the SOGS (Strong, Lesieur, Breen, Stinchfield, & Lejuez, 2004), which winnowed the original 20 items to 6 items that provide excellent reliability and validity by reflecting divergent levels of severity. The count of positive responses to these six items represents the third alternative scoring of SOGS, the IRT approach. The comparison between this method and methods based on symptom matching is especially interesting because the six IRT-identified items provide notably narrow coverage of the *DSM-IV* symptoms. Five of the items (14, 16a, 16d, 16g, and 16i) reflect borrowing money as a result of gambling, and the sixth (Item 7) reflects increasing amount gambled (under permissive symptom matching, but not under conservative symptom matching).

The Present Study

We examined the performance of the SOGS in a relatively large sample of community-based frequent gamblers, defined as self-reported gambling at least weekly. The focus on frequent gamblers excluded individuals for whom the SOGS would be largely irrelevant and permitted the enrollment of substantial proportions of individuals at various levels of clinical severity. The utility of the SOGS was evaluated in predicting a widely used diagnostic interview, the Structured Clinical Interview for Pathological Gambling (SCI-PG; Grant, Steinberg, Kim, Rounsaville, & Potenza, 2004), and its convergent validity was assessed by comparing its correlations with measures of several personality traits, mood, and addictive behaviors with those manifested by the dimensional ratings on the SCI-PG. The study evaluated the SOGS both dimensionally and categorically and explored the utility of alternative scoring regimens.

Method

Participants were recruited using flyers, newspaper advertisements, and word of mouth from the community. Eligibility criteria were (a) adult (i.e., age 18+), (b) frequent gambling (i.e., at least weekly), (c) adequate literacy (i.e., 9+ grade education), and (d) self-reported ability to use a computer. A total of 368 individuals enrolled and were compensated \$30 for 3 hours of participation in a larger project. Fifteen individuals were excluded for missing data and very low effort, defined as meeting two of the following three criteria: lowest 5% of scores on the Shipley test of cognitive functioning, less than 88% consistent choices in the Monetary Choice Questionnaire, and subjective identification by a cognizant research assistant. This resulted in a total sample of 353 participants (78% male; age $M = 35.6$ [$SD = 12.4$]; 52% Caucasian, 43% African American). The sample was lower in income than the population as a whole, with 52% reporting income <\$15,000 per year and another 24% reporting \$15,000 to \$30,000.

In addition to the SOGS ($M = 10.5$; $SD = 4.96$; Cronbach's $\alpha = .87$), the following instruments were administered. To assess gambling symptoms, two instruments were used. The SCI-PG (Grant et al., 2004) is a semistructured clinical interview assessing *DSM-IV*

symptoms of PG over the past year ($M = 3.23$, $SD = 2.84$, $\alpha = .86$). Categorically, 23% of the sample exhibited no symptoms, 46% exhibited 1 to 4 symptoms, and 31% exhibited 5 or more symptoms, meeting diagnostic criteria for PG. The SCI-PG was administered by MS-level research assistants, who were trained by licensed clinical psychologists (JM, JDM). Symptom counts were used as a continuous index of PG severity.

PG is robustly associated with co-occurring substance use problems and also with mood-related personality traits, most notably negative affect (MacLaren, Fugelsang, Harrigan, & Dixon, 2011; Slutske, Caspi, Moffitt, & Poulton, 2005). Consequently, to facilitate assessment of the SOGS's convergent validity with regard to personality, mood, and addictive behaviors, the following instruments were also administered. The Positive and Negative Affect Schedule–Expanded Form (PANAS-X; Watson & Clark, 1994) is a 60-item self-report measure of affect. We report on the factors of positive affect (10 items; $\alpha = .88$) and negative affect (10 items; $\alpha = .90$). The Multidimensional Personality Questionnaire–Brief Format (MPQ-BF; Patrick, Curtin, & Tellegen, 2002) is a 155-item, self-report inventory that assesses 11 personality traits and 3 domains (i.e., positive emotionality, negative emotionality, and constraint) from Tellegen and Waller's (2008) model. Alphas ranged from .65 (harm avoidance) to .87 (stress reaction) for the 11 traits and from .79 (constraint) to .91 (negative emotionality) for the 3 domains. The Fagerstrom Test for Nicotine Dependence (FTND; Heatherton, Kozlowski, Frecker, & Fagerstrom, 1991) is a 6-item validated measure of nicotine dependence. The **Alcohol** Use Disorders Identification Test (AUDIT; Saunders, Aasland, Babor, de la Fuente, & Grant, 1993) is a 10-item assessment of alcohol consumption and negative consequences. Total scores range from 0 to 40, with a score of 8 or higher indicating hazardous drinking ($\alpha = .87$).

African Americans have been reported to have systematically more severe gambling problems (Petry, Stinson, & Grant, 2005). Because our sample included robust numbers of both White and African American participants, we tested for differences between racial subsamples, to ascertain whether SOGS performed equivalently for both groups. Participants of other races were excluded from these analyses due to small subsample sizes.

Results

At the dimensional level, SOGS manifested a strong correlation with interview ratings of *DSM-IV* PG symptoms ($r = .66$). These aspects of SOGS's performance were consistent across racial categories. For the White subsample, $\alpha = .88$ and $r = .67$ ($p < .001$). For the African American subsample, $\alpha = .87$ and $r = .60$ ($p < .001$). Neither gender nor race moderated the relations between SOGS and *DSM-IV* PG symptoms. SOGS correlated virtually as strongly with SCI-PG ratings of *DSM-5* PG symptoms ($r = .66$ for White subsample, $r = .60$ for African American subsample).

We next examined the relationships manifested by the gambling scores with demographic variables, affect, personality, and substance abuse, which are shown in Table 1. To test the similarity of the correlations manifested by the interview-based PG symptoms and the SOGS, we (a) tested whether the correlations were significantly different (test of dependent r s) with each criterion and (b) examined the similarity of the overall sets of correlations

using a double-entry q -correlation (McCrae, 2008). In general, both sets of gambling scores were correlated with race, negative affect, trait negative emotionality, and substance dependence. Across 20 comparisons, there were no statistically significant differences between the correlations manifested by the SCI-PG and the SOGS with the 20 external criteria. The two sets of correlations were extremely similar, as indicated by an intraclass correlation of .95.

At the categorical level, however, relating SOGS designations (score ≥ 5 indicating PPG) to *DSM-IV* designations of PG (symptom count ≥ 5 indicating PG), the results were less favorable. Table 2 presents the cross-tabulations of SCI-PG and SOGS scores, in quadrants reflecting true positives, true negatives, false positives, and false negatives. In a sample with a moderate base rate of PG (32.1%), 195 false positives were observed versus only 1 false negative. The FAR is $195/(195 + 47) = 81\%$, and the FDR is $195/(195 + 110) = 64\%$. In other words, 64% of SOGS's PPG designations are incorrect, and 81% of individuals without PG were incorrectly judged to be PPG, using the SOGS.

Table 3 presents the same cross-tabulations, with SCI-PG rescored and the table reconfigured to reflect the diagnostic criteria changes in *DSM-5*. Symptom 8 (illegal acts) was removed from each participant's score, and scores of 4 are designated as PG^+ . Consequently, the percentage of participants with no symptoms remained 23%. A total of 37% showed 1 to 3 symptoms (compared with 46% with 1–4 symptoms under *DSM-IV* criteria) and 39% exhibited 4 to 9 symptoms (compared with 31% showing 5–10 symptoms under *DSM-IV* criteria). At a dimensional level, SOGS scores are an equally strong predictor of *DSM-5* symptom count as of *DSM-IV* symptoms ($r = .66$). The changes in diagnostic criteria from *DSM-IV* to *DSM-5* move some who were previously considered to have subclinical symptoms to PG^+ , with correction for the symptom of committing illegal acts, whereas the designation of those with no symptoms remains unchanged. Although the reduction in criterion symptom count diminishes the number of false positives by 16% (from 195% to 163%), the FAR and FDR are relatively little changed, at 78% and 53%, respectively. Clearly, there remains a major disparity between the number of false positives (163) and false negatives (1).

Figure 1 depicts the FAR, PPV, sensitivity, and specificity for all possible cut points from 1 to 20 within this sample, using *DSM-IV* criteria. If the cut point were changed from 5 to 8, 53 individuals' designations would be changed from false positives to true negatives, and only one would be changed from a true positive to a false negative. The FAR would decline from 81% to 59%, and the PPV would increase from 36% to 43%. Sensitivity would remain in excess of 98%. Nearly half of those who would no longer be designated as PPGs (25 out of 53) have no *DSM* symptoms. The cut point would need to be increased at least to 10 in order for the PPV to exceed 0.5. At this point, more than half of those designated as PPG would warrant a PG diagnosis, and the designation of "probable pathological gambler" would be accurate within this sample. The cut point would need to be increased to 13 in order for sensitivity to be exceeded by specificity and to 14 for false positives to be exceeded by false negatives.

We conducted similar analyses with racial subgroups. Sensitivity and specificity were generally similar in White and African American subsamples, although both were lower among African Americans. For example, at a cut point of 5, both groups showed high sensitivity (1.00 among Whites, .94 among African Americans) but extremely poor specificity (0.19 among Whites and 0.08 among African Americans). At a cut point of 13, sensitivity was 0.64 among Whites and 0.61 among African Americans. Specificity was 0.71 among Whites and 0.56 among African Americans. The lowest cut point at which specificity exceeded sensitivity was 13 for Whites and 14 for African Americans.

Figure 2 depicts the same diagnostic quality metrics as Figure 1, incorporating *DSM-5* criteria. At a cut point of 8, 51 individuals' designations would be changed from false positives to true negatives, and only three would be changed from a true positive to a false negative. The FAR would decline from 78% to 53%, and the PPV would increase from 45% to 54%. Sensitivity would remain in excess of 97%. The cut point would need to be increased at least to 7 in order for the PPV to exceed 0.5, to 12 in order for sensitivity to be exceeded by specificity, and to 13 for false positives to be exceeded by false negatives.

Alternative Scoring Methods

Using the conservative symptom matching approach, the set of four symptoms identified by Stinchfield (2002) as measured by SOGS resulted in poor reliability ($\alpha = .58$) and a much smaller correlation with SCI-PG scores ($r = .20, p < .01$). Using the permissive symptom matching approach increased the reliability of the symptom-matching approach ($\alpha = .72$) but did not improve its correlation with SCI-PG scores ($r = .19, p < .01$). Using IRT-informed scoring, the set of symptoms identified by Strong et al. (2004) as having the strongest psychometric properties manifested relatively low internal reliability, $\alpha = .66$, and a small correlation with SCI-PG scores ($r = .20, p < .01$). Because the alternative scoring methods reflect a continuous rather than a categorical approach, and none of the methods incorporated illegal acts as a symptom, there is no distinction to be made in these analyses between *DSM-IV* and *DSM-5* criteria.

Discussion

In a large, diverse community sample of frequent gamblers, the SOGS generally performed well when used dimensionally, but exhibited serious limitations when used categorically on the basis of the traditional cut point of 5. Attempts to rescore the SOGS, either to align it more closely with *DSM* PG criteria or to reflect items with previously demonstrated diagnostic power, did not improve classification properties of the measure as all three alternative scoring strategies manifested substantially smaller correlations with PG symptoms. These limitations remain substantially unchanged when *DSM-5* criteria are applied.

Replicating many previous studies, the SOGS manifested good internal consistency, a strong correlation with symptoms from a widely used *DSM-IV* clinical interview, and an almost identical pattern of convergent validity correlations as those manifested by the interview ratings of *DSM-IV* PG symptoms. These correlational results, combined with similar previously reported correlations (e.g., Cox, Enns, & Michaud, 2004; Stinchfield, 2002),

suggest that the dimensional SOGS score serves well as a research screen, albeit not as a reliable tool for individual diagnosis. Hence, for the numerous studies in which SOGS has been used dimensionally and in a research setting (i.e., establishing correlational relationships between other variables of interest and SOGS scores reflecting degree of gambling pathology), the current analyses present no challenge, despite the mismatch between SOGS items and modern *DSM* criteria. We also found that these findings extend to the important population of African Americans (Petry et al., 2005).

When the SOGS is used in a categorical manner, there are problems due to the presence of vastly more false positives than false negatives. We are not the first to discuss this possibility, although few sources present independent data to support this supposition (Stinchfield, 2002, being a notable exception). The current study adds to this data set, using a single, community-based sample where both Stinchfield (2002) and Lesieur and Blume (1987) relied on multiple samples to achieve robust representation in all cells.

In clinical applications, false positives are arguably less severe errors than false negatives (Stinchfield, 2002), as an untreated condition may cause more significant problems than the cost, side effects, and other consequences of a false positive diagnosis. Furthermore, when an instrument is used as a clinical screen—that is, a resource-non-intensive instrument to identify a smaller pool of at-risk individuals to be subjected to further, more costly testing—it is particularly appropriate to tolerate a large number of false positives to avoid excluding individuals in need of help as false negatives. However, the SOGS has been widely used beyond this narrow designation as a clinical screen. Where the SOGS has been used to establish pseudo-experimental groups, the inclusion of some NPGs in PG groups could have various adverse impacts on the validity of studies. In the numerous prevalence studies that use SOGS as a primary indicator of PG status, a high false positive rate would systematically inflate prevalence rates. If the present extreme imbalance between false positive and false negatives is representative of the SOGS's performance in subclinical populations in general, the inflation in prevalence rates may be substantial.

The data in this study, where false positives vastly outnumber false negatives, argue for the need to increase the SOGS cut point to reduce false positives, without inordinately increasing false negatives. In the current data set, the traditional cut point of 5 correctly identified more than 99% of PG cases (sensitivity) but also identified more than 80% of NPGs incorrectly as PPG (<20% specificity), rendering the designation of “probable pathological gambling” as simply incorrect within this sample. Increasing the cut point would have a multifaceted impact, with its utility depending on the underlying purpose for using the SOGS. For a clinical screen, many false positives should be tolerated, under the assumption that they will be eliminated at later testing. Although the cut point of 5 would not have eliminated a large proportion of our participants from further testing (45 out of 336 or 13%), this is likely a consequence of the choice to recruit frequent gamblers. A higher proportion of individuals would be excluded from further testing within a general population, who do not gamble or gamble infrequently. These individuals appear unlikely to be experiencing PG but missed in a screen with a higher cut point.

If SOGS is used diagnostically, despite the clear divergence of its items from modern *DSM* symptoms, then specificity should be maximized while maintaining high sensitivity. For prevalence studies, the cut point should make the number of false positives equal the number of false negatives. As depicted in Figure 1, increasing the cut point from 5 to 8 more than doubles specificity (from .19 to .41) while maintaining sensitivity in excess of .98. Even increasing the cut point to 11 maintains sensitivity greater than .90, while increasing specificity by more than an additional .25, to .67. For the purpose of having false positives approximately equal to false negatives, the cut point should be raised even further, to 12 or 13. Finally, for the designation PPG to reflect that more individuals so designated are PG than are NPG, the cut point would need to be set to 10 or higher.

As the *DSM-5* reduces the diagnostic criterion from 5 symptoms to 4, one would expect the problem of false positives to become less severe. (The change will render more individuals' gambling problems as PG⁺; therefore, fewer SOGS⁺ designations will be falsely positive.) Indeed, the results of the current study show a less extreme imbalance between false positives and false negatives, although reducing the ratio from 195:1 to 163:1 scarcely achieves parity. These data suggest that a substantial change in the cut point should be adopted for categorical use of the SOGS, even under *DSM-5* criteria. We recommend a cut point of 10 for use with *DSM-IV-TR* criteria, as it retains high sensitivity while achieving more acceptable levels of specificity, achieving a false alarm rate less than 50%, conferring greater validity to the designation of probable pathological gambler, and using the smallest possible change in the cut point to achieve these goals.

For implementation of *DSM-5* criteria, we advocate a two-pronged approach. As an authentic clinical screen, a relatively permissive cut point of 8 has the properties listed above, and maintains sensitivity in excess of 97%, which merits deference in clinical screens. For research purposes that evaluate prevalence or assign to pseudo-experimental groups, it is more important to minimize overall errors, and to balance false positive and false negative rates. The number of errors (combined false positives and false negatives) is minimized at a cut point of 11 (80 errors). The difference between sensitivity and specificity is minimized at a cut point of 12 (both having values of .76), and the number of errors with a cut point of 12 is only marginally greater than the minimum (83). We therefore advocate a score of 12 as the preferred SOGS cut point for research purposes with *DSM-5* criteria.

A recent literature argues for disordered gambling to be considered dimensionally on a continuum rather than in dichotomous (PG vs. NPG) or trichotomous (PG, NPG, and problem gambler) categories, which would render the issue of cut point as moot. Slutske et al. (2011) maintain that a critique based on high false positive rates “is not an especially damning criticism because it is primarily a function of the cutoffs used to make a diagnosis ... and this has been changed with every revision of the *DSM*, often without empirical justification” (p. 744). We maintain that it is important to distinguish between dimensional and categorical uses of SOGS. In the first case, we conclude (in concert with Slutske et al., 2011; Strong et al., 2004) that the SOGS remains a generally useful instrument when analyzed dimensionally. However, there remain substantial research and clinical domains that continue to use the SOGS in a categorical way (e.g., Volberg, 1996); in these endeavors, it is clear that a different cut point is necessary.

Limitations

The present study sampled only regular gamblers, and although it included robust numbers of NPGs, it excluded large swaths of the general population that gamble infrequently or do not gamble at all. Studies that include the entirety of the general population would be expected to reveal much larger true negative cell totals, reflecting individuals with subthreshold scores on both SOGS and *DSM-IV* criteria and potentially reducing error rates. On the other hand, classification of nongamblers and highly infrequent gamblers is not the object of the SOGS and the point of studies like the current one is to clarify its classification ability among individuals with relevant levels of the behavior in question. Of note, the present sample also reflected lower income distributions than are observed in the population at large. While lower income individuals are an important subpopulation to study, results of the current study could differ from characteristics of the general population due to income considerations.

Conclusions

For uses in which the SOGS is utilized dimensionally, which have encompassed much research of value, the current results reaffirm its utility with both *DSM-IV* and *DSM-5* criteria. However, for endeavors in which the SOGS is used categorically, including prevalence studies, studies in which it is used as the sole diagnostic instrument of PG, and research that assigns to pseudo-experimental groups on the basis of diagnostic categories, the problem of frequent and imbalanced false positives identified by the SOGS are substantial. They are diminished but still evident under *DSM-5* criteria, compared with *DSM-IV* criteria. The current data suggest that a revised cutoff score of 10 may be more useful with *DSM-IV* criteria; with *DSM-5* criteria, a cut point of 8 is advocated for clinical screens, and a cut point of 12 is advocated for prevalence and pseudo-experimental studies.

Acknowledgments

We are very grateful for the contributions of the research assistants in the Georgia Decision Laboratory.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported in part by grants from the National Center for Responsible Gaming (ASG, JDM, JM, CEL, WKC) and the National Institutes of Health (K23 AA016936—JM; P30 DA027827—ASG, JM).

References

- Cox BJ, Enns MW, Michaud V. Comparisons between the South Oaks Gambling Screen and a *DSM-IV*-based interview in a community survey of problem gambling. *Canadian Journal of Psychiatry*. 2004; 49:258–264.
- Duvarci I, Varan A, Coskunol H, Ersoy MA. *DSM-IV* and the South Oaks Gambling Screen: Diagnosing and assessing pathological gambling in Turkey. *Journal of Gambling Studies*. 1997; 13:193–206. [PubMed: 12913386]
- Grant JE, Steinberg MA, Kim SW, Rounsaville BJ, Potenza MN. Preliminary validity and reliability testing of a structured clinical interview for pathological gambling. *Psychiatry Research*. 2004; 128:79–88. [PubMed: 15450917]
- Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerstrom Test for Nicotine Dependence: A revision of the Fagerstrom Tolerance Questionnaire. *Journal of Addiction*. 1991; 86:119–127.

- Lesieur HR, Blume SB. The South Oaks Gambling Screen (SOGS): A new instrument for the identification of pathological gamblers. *American Journal of Psychiatry*. 1987; 144:1184–1188. [PubMed: 3631315]
- MacLaren VV, Fugelsang JA, Harrigan KA, Dixon MJ. The personality of pathological gamblers: A meta-analysis. *Clinical Psychology Review*. 2011; 31:1057–1067. [PubMed: 21802620]
- McCrae RR. A note on some measures of profile agreement. *Journal of Personality Assessment*. 2008; 90:105–109. [PubMed: 18444102]
- Miller JD, MacKillop J, Fortune EE, Maples J, Lance CE, Campbell WK, Goodie AS. Personality correlates of pathological gambling derived from Big Three and Big Five personality models. *Psychiatry Research*. 2013; 206:50–55. [PubMed: 23078872]
- Patrick CJ, Curtin JJ, Tellegen A. Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment*. 2002; 14:150–163. [PubMed: 12056077]
- Petry, NM. *Pathological gambling: Etiology, comorbidity, and treatment*. Washington, DC: American Psychological Association; 2005.
- Petry NM, Stinson FS, Grant BF. 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*. 2005; 66:564–574. [PubMed: 15889941]
- Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction*. 1993; 88:791–804. [PubMed: 8329970]
- Slutske WS, Caspi A, Moffitt TE, Poulton R. Personality and problem gambling: A prospective study of a birth cohort of young adults. *Archives of General Psychiatry*. 2005; 62:769–775. [PubMed: 15997018]
- Slutske WS, Zhu G, Meier MH, Martin NG. Disordered gambling as defined by the Diagnostic and Statistical Manual of Mental Disorders and the South Oaks Gambling Screen: Evidence for a common etiologic structure. *Journal of Abnormal Psychology*. 2011; 120:743–751. [PubMed: 21443290]
- Stinchfield R. Reliability, validity, and classification accuracy of the South Oaks Gambling Screen (SOGS). *Addictive Behaviors*. 2002; 27:1–19. [PubMed: 11800216]
- Strong DR, Lesieur HR, Breen HR, Stinchfield R, Lejuez CW. Using a Rasch model to examine the utility of the South Oaks Gambling Screen across clinical and community samples. *Addictive Behaviors*. 2004; 29:465–481. [PubMed: 15050667]
- Tellegen, A.; Waller, N. Exploring personality through test construction: Development of the multidimensional personality questionnaire. In: Boyle, G.; Matthews, G.; Saklofske, D., editors. *The SAGE handbook of personality theory and assessment: Volume 2 — Personality measurement and testing*. London: Sage; 2008. p. 261-293.
- Volberg RA. Prevalence studies of problem gambling in the United States. *Journal of Gambling Studies*. 1996; 12:111–128. [PubMed: 24233911]
- Watson, D.; Clark, LA. *The PANAS-X: Manual for the Positive and Negative Affect Schedule—Expanded Form*. University of Iowa; Iowa City: 1994. Unpublished manuscript
- Wickwire EM, Burke RS, Brown SA, Parker JD, May RK. Psychometric evaluation of the National Opinion Research Center *DSM-IV* Screen for Gambling Problems (NODS). *American Journal on Addictions*. 2008; 17:392–395. [PubMed: 18770081]

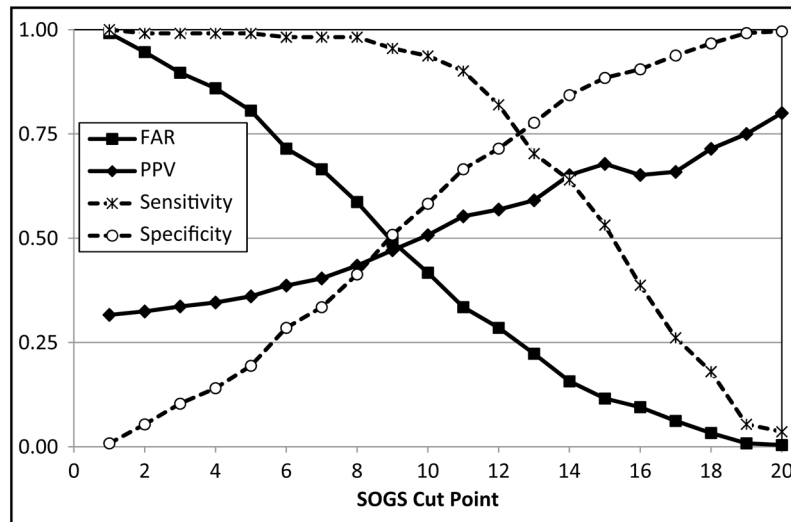


Figure 1. Impact of hypothetical cut points on false alarm rate (FAR), positive predictive value (PPV), sensitivity, and specificity with regard to *DSM-IV* criteria.

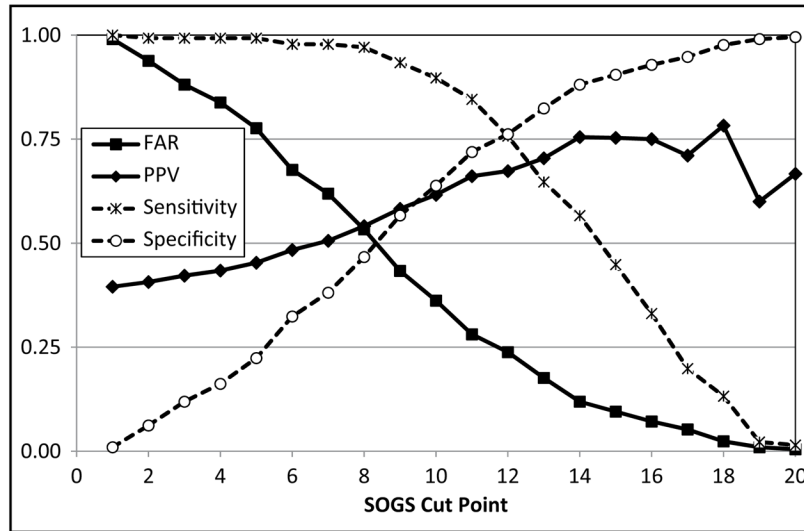


Figure 2. Impact of hypothetical cut points on false alarm rate (FAR), positive predictive value (PPV), sensitivity, and specificity with regard to *DSM-5* criteria.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Convergent Validity of the SCI-PG and SOGS Measures of Gambling.

	SCI-PG	SOGS
Demographics		
Gender	.17*	.12
Race	.21*	.21*
Affect		
Positive	-.09	-.11
Negative	.26*	.28*
Personality		
Positive Emotionality	-.12	-.06
Well-Being	-.14	-.13
Social Potency	-.04	.00
Achievement	-.01	.04
Social Closeness	-.15	-.08
Negative Emotionality	.27*	.28*
Stress Reaction	.29*	.29*
Alienation	.26*	.31*
Aggression	.07	.07
Constraint	.03	.09
Control	-.13	-.08
Harm Avoidance	.03	.07
Traditionalism	.20*	.23*
Absorption	.00	-.02
Substance use		
Tobacco Dependence	.17*	.28*
Alcohol Dependence	.16*	.19*
r_{ICC}	.95*	

Note. SCI-PG = Structured Clinical Interview for Pathological Gambling; SOGS = South Oaks Gambling Screen; MPQ = Multidimensional Personality Questionnaire-Brief Format. Gender coded as 0 = *men*, 1 = *women*; Race coded as 0 = *Caucasian*, 1 = *African American*, other races excluded. Correlations between SCI-PG symptoms and MPQ are originally reported in Miller et al. (2013).

* $p < .01$.

Table 2
Cross-Tabulation of DSM-IV Symptom Counts and Diagnoses With SOGS Scores and Classifications.

SOGS Score	DSM-IV Score										PG ⁺		
	0	1	2	3	4	PG ⁻	5	6	7	8		9	10
0	2	0	0	0	0	0	0	0	0	0	0	0	0
1	8	0	1	0	0	1	0	0	0	0	0	0	0
2	8	3	0	0	0	0	0	0	0	0	0	0	0
3	7	2	0	0	0	0	0	0	0	0	0	0	0
4	5	5	2	1	0	0	0	0	0	0	0	0	0
SOGS ⁻	True negative				44	False negative				1			
5	8	2	6	1	1	1	0	0	0	0	0	0	0
6	6	4	2	0	0	0	0	0	0	0	0	0	0
7	9	4	1	3	1	0	0	0	0	0	0	0	0
8	7	5	4	4	2	1	2	0	0	0	0	0	0
9	2	5	5	3	2	1	0	1	0	0	0	0	0
10	3	2	6	6	2	1	1	1	1	0	0	0	0
11	1	3	3	2	3	1	1	5	2	0	0	0	0
12	1	2	5	4	2	5	4	0	2	2	0	0	0
13	2	3	2	5	4	4	1	0	2	0	0	0	0
14	0	0	3	2	5	4	2	2	1	1	1	1	1
15	2	1	0	2	0	4	2	5	4	0	0	0	0
16	1	0	0	3	3	2	3	4	2	3	0	0	0
17	3	1	1	1	1	3	1	0	2	2	1	1	1
18	0	0	1	1	3	1	2	2	6	0	0	2	2
19	1	0	0	0	0	0	0	1	0	0	0	1	1
20	0	0	1	0	0	0	1	0	1	0	0	2	2
SOGS ⁺	False positive				184	True positive				107			

Note. SOGS = South Oaks Gambling Screen.

Table 3
 Cross-Tabulation of DSM-5 Symptom Counts and Diagnoses With SOGS Scores and Classifications.

SOGS Score	DSM-5 Score										PG ⁺	
	0	1	2	3	PG ⁻	4	5	6	7	8		9
0	2	0	0	0	0	0	0	0	0	0	0	0
1	10	0	1	0	0	1	0	0	0	0	0	0
2	9	3	0	0	0	0	0	0	0	0	0	0
3	7	2	0	0	0	0	0	0	0	0	0	0
4	5	5	2	1	0	0	0	0	0	0	0	0
SOGS ⁻	True negative			47	False negative			1				
5	9	4	7	1	1	1	0	0	0	0	0	0
6	6	4	2	0	0	0	0	0	0	0	0	0
7	10	4	1	3	1	0	0	0	0	0	0	0
8	7	5	4	5	2	1	2	0	0	0	0	0
9	2	5	5	3	3	1	0	1	0	0	0	0
10	3	2	6	6	3	1	0	2	1	0	0	0
11	1	3	3	2	3	1	1	5	2	0	0	0
12	1	2	5	5	2	5	4	0	2	2	2	2
13	2	3	2	5	4	4	1	0	2	0	0	0
14	0	0	3	2	5	4	2	3	1	1	1	1
15	2	1	0	2	0	4	2	5	5	0	0	0
16	1	0	0	3	4	2	3	4	2	3	3	3
17	3	1	1	1	1	3	1	0	2	2	2	2
18	0	0	1	2	3	1	3	2	6	0	0	0
19	1	0	0	0	0	0	0	1	0	0	0	0
20	0	0	1	0	0	0	1	0	1	0	1	0
SOGS ⁺	False positive			163	True positive			135				

Note. SOGS = South Oaks Gambling Screen.