Journal Information
Journal ID (publisher-id): jgi
ISSN: 1910-7595
Publisher: Centre for Addiction and Mental Health

Article Information

Publication date: September 2005
Publisher Id: jgi.2005.14.4
DOI: 10.4309/jgi.2005.14.4

Interpreting prevalence estimates of pathological gambling: Implications for policy

Blasé Gambino

Affiliation: Boston, Massachusetts, U.S.A. E-mail: blasegambinophd@aol.com

[This article prints out to about 19 pages.]

This article was peer-reviewed. Submitted: August 8, 2003. All URLs cited were available at the time of submission. Accepted: March 14, 2005.

For correspondence: Blasé Gambino, PhD, 10 Ellet Street, Boston, MA 02122 U.S.A., Telephone: 617-282-2560, e-mail: blasegambinophd@aol.com
Competing interests and funding: None declared.

Blasé Gambino (PhD, experimental psychology, University of Massachusetts at Amherst) currently serves on the editorial review board of the *Journal of Gambling Studies*. His publications include articles on mathematical models to eliminate bias in prevalence estimation research for single- and two-stage prevalence designs, as well as articles on the epidemiology of pathological gambling. Co-editor of *Compulsive Gambling: Theory, Research and Practice* (1989). His current interests are in the epidemiology of pathological gambling, latent structure, and taxometrics.

Abstract

Some guidelines for interpreting prevalence estimates for the purpose of establishing the number of pathological gamblers in the community are presented. The analysis is based on the concept of the likelihood ratio, a recommended procedure for validating criteria for defining cases based on test scores. It is shown that the likelihood ratio can be employed with available estimates of prevalence to translate cut-off scores into positive predictive value. Those cut-off scores associated with high positive predictive values provide an empirical measure of confidence that those gamblers who meet or exceed the cut-off criterion are pathological gamblers. A potential limitation of the analysis is the possible

specificity of results to the validation studies employed to compute likelihood ratios and to the specific estimates of prevalence used to determine positive predictive value. A recommendation is presented for obtaining study- or community-specific validation evidence.

Introduction

Prevalence estimates of pathological gambling reflect choices and assumptions made by researchers (Gambino, 1997a). Choices include decision rules, such as the cut-off point used to define a case (Brenner & Gefeller, 1997; Dickerson et al., 1996) and the time period over which cases are to be defined, for example, six-months (Abbott & Volberg, 1991), past-year (Welte et al., 2001), or lifetime (Volberg, 1994). Assumptions include our confidence in the validity of the measurement systems that are employed to obtain estimates (e.g., Gambino, 1999a; Stinchfield, 2002, 2003).

On the surface, determining how many pathological gamblers there are in a community seems to be a straightforward task. Define who is or is not a pathological gambler, apply this “working” definition to a representative sample drawn from the population of interest, count how many meet the chosen definition, and divide by the number of eligible respondents. In practice the issue is complicated by the lack of consensus over the most appropriate means of defining a “case” in community surveys on pathological gambling (Abbott & Volberg, 1999; Dickerson, 1993; Dickerson & Volberg, 1996; Dickerson et al., 1996; Gambino, 1997a, 1999a; Poulin, 2002; Shaffer et al., 1997; Walker & Dickerson, 1996). An additional complication is the definition of “eligible respondents”; should non-gamblers be counted in that number or not (Shaffer et al., 1997)?

It has been observed that case-definition strategies are the “sine qua non” for most epidemiologic research (Zahner et al., 1995, p. 23). In the absence of a case-definition, the relevant events or states cannot be identified and counted, and prevalence or other measures of interest cannot be obtained. Agreement on some form of classification always entails some degree of arbitrariness. The convenience of using shared case-definitions to assign individuals into categories as cases and non-cases is fundamental to communication among researchers and clinicians (Rose & Barker, 1978). Its utility stems from the achievement of comparability among data sources and researchers; and, in addition, it permits the testing of etiologic and other hypotheses. Shared case-definitions also have implications for communicating with policy makers. A major task for researchers will be how to “calculate” and effectively “communicate” the implications of their findings, including the meaning of agreed-upon case-definitions to policy makers (Koplan et al., 1999, p. 1153).

A complicating factor in the interpretation of prevalence estimates is the lack of agreement on how to deal with the occurrence of diagnostic errors (Abbott & Volberg, 1999; Gambino, 1999a; Shaffer & Korn, 2002; Volberg, 1999). Given the expected lack of perfect discriminability of any definition (Kraemer, 1992; Zhou et al., 2002), any group classified as cases (positive test outcomes) will include some non-disordered individuals (false positives), and any group classified as non-cases (negative test outcomes) will include some who are truly disordered (false negatives). The basic question in the case of prevalence estimation is a simple one. Given the presence of errors, are sample prevalence estimates biased or unbiased (Shaffer et al., 1997)? Bias refers to whether sample estimates tend, on average, to overestimate (positive bias) or underestimate (negative bias) the true population prevalence (Gambino, 1997b).

Many of the proposed solutions are complicated and generally entail mathematical and statistical models (Gambino, 1997b, 1999a, 1999b; Garrett et al., 2002; Hui & Walter, 1980; Rogan & Gladen, 1978; Staquet et al., 1981). Available solutions have seldom been employed in studies of pathological gambling, although this is true for other medical and psychiatric disorders as well (Faraone & Tsuang, 1994). The failure to apply these procedures is generally conceded to be the perception that these models are viewed as too mathematically complex. The general focus of these models has been on obtaining precise estimates of error rates, although that is a simplification. An alternative solution is presented below.

Terms and definitions

There are four possible outcomes from testing a sample of respondents drawn from a specific population for the purposes of assessing the presence or absence of pathological gambling. These are presented in Table 1, where a, b, c, and d are, by convention, labeled as true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN), respectively. The terms positive (predicting presence of the disorder) and negative (predicting absence of the disorder) simply mean that the respondent met or did not meet the criterion for defining a case.

Table 1 also presents four measures of diagnostic accuracy. Diagnostic accuracy may be defined as the ability of a test to discriminate those with the disorder from those in whom the disorder is absent (Zhou et al., 2002). The four measures of diagnostic accuracy presented in Table 1 may be further distinguished by the labels test accuracy and predictive accuracy. Test accuracy is represented by sensitivity, defined as the proportion of those with the disorder with positive test results (true positives); and specificity, the proportion of those without the disorder with negative test results (true negatives). Predictive accuracy is represented by positive predictive value, the proportion of positive tests that are true positives, and negative predictive value, the proportion of negative tests that are true negatives.

These four measures are related but not identical since they are based on different denominators (Table 1). The primary distinction lies in the fact that sensitivity and specificity are independent of population prevalence, whereas positive and negative predictive value will change as a function of prevalence. In general, the positive (negative) predictive value of any instrument will be high (low) when applied to populations with high (low) prevalence rates. As prevalence decreases positive (negative) predictive value will decrease (increase). Technically, predictive values are known as specific rates since these are specific to the prevalence of the population being tested as well as the sensitivity and specificity of the test instrument employed.

Defining a useful case-definition

The minimal requirement for a suitable diagnostic or screening case-definition (Meehl & Rosen, 1955) is that it yields a higher percentage of positive findings among the truly disordered (its sensitivity) than among the truly non-disordered (its lack of specificity). Put simply, sensitivity (Se), the true positive rate of the test among pathological gamblers, must be greater than 1 – specificity (1 – Sp), the false positive rate of the test among non- pathological gamblers. The above requirement also implies each of the following relationships: Se + Sp > 1 and PPV > P where PPV = positive predictive value and P = true prevalence; NPV > 1 - P where NPV = negative predictive value and Se > Pp where Pp = the sample prevalence estimator (the observed proportion of positive outcomes identified as those respondents meeting criteria for caseness).

Current conventions for defining a case. The two most frequently employed instruments for conducting research on pathological gambling (Shaffer et al., 1997) are the SOGS (South Oaks Gambling Screen) (Lesieur & Blume, 1987), a 20-item instrument, and the current clinical definition accepted by the American Psychiatric Association, the DSM-IV, a 10-item test (APA, 1994). The general convention for defining a case of pathological gambling adopted for both instruments is that those who respond positive to five or more of the clinical indicators incorporated in the instrument will be defined as cases of pathological gamblers. Those individuals who score less than five will be defined as not being cases of pathological gamblers. Scores that are less than five, but greater than zero, i.e., 1–4, have been given a variety of labels including problem, potential pathological, at-risk, and level-two gamblers (National Research Council, 1999). This varied nomenclature is another source of confusion in the literature on the estimation of prevalence (Poulin, 2002).

A second source of confusion is related to the number of items on the two instruments. The number of items on the SOGS relative to DSM-IV provides 10 additional opportunities to meet the recommended criterion of five or higher and, in

part, may explain the higher levels of prevalence reported for the SOGS (Shaffer et al., 1997). A third source of difficulty flows from the expressions of dissatisfaction with both instruments. The net result of this dissatisfaction has been a continuing effort to develop and validate alternative instruments, mostly in the form of variants on both the SOGS and the DSM-IV (Shaffer & Korn, 2002).

In practice, some investigators have argued for a different cut-off point for defining a case. For example, Dickerson et al. (1996) have argued that for the SOGS the criterion should be set at 10 to reflect the average scores obtained from gamblers in treatment. Stinchfield (2003), employing discriminant analysis, a statistical procedure for separating those with from those without the disorder, has argued that a criterion score on DSM-IV of four or higher is a more accurate discriminator between the presence and absence of pathological gambling than the recommended criterion of five. What are the implications of raising or lowering criteria relative to the recommended convention of a criterion score of five or higher?

Setting the criterion bar for case ascertainment. If it is important to protect against false positives, the researcher may set stringent criteria, e.g., eight or higher, but this comes at the cost of an increased likelihood of false negatives. Protecting against false negatives by the use of less stringent criteria, e.g., three or higher, has the opposite effect. Now it is unlikely that many cases will be missed, but there is an increased probability that many of the presumptive diagnoses will represent false positives. The first strategy (stringent criteria) provides conservative estimates of prevalence; the latter strategy (less stringent criteria) results in liberal estimates (Gambino, 1997a). An important implication of raising and lowering the cut-off point is often overlooked. The use of a cut-off score to separate individuals into two categories, pathological or not pathological, is always arbitrary, as is the implication that the disorder is dichotomous in nature. All that can be stated as factual is that once a cut-off has been set, then the following must be true: Those who score at or above the criterion can *only* be true positives or false positives. Those who score below the criterion can *only* be true negatives or false negatives. In practice, since pathological gambling is a construct and not in the realm of public scrutiny, the truth or falsity of these four labels can never be known with complete certainty.

Protecting against false positives, or against false negatives? The decision to protect against false negative or false positive errors will be conditional on the goals of the decision-maker, and the severity of the consequences of making false positive or false negative errors. In the clinical setting, for example, the test outcome is not the sole source of evidence. A detailed history of the gambler is usually taken in addition to the application of one or more tests. The clinician is usually more concerned with false negatives than false positives. The clinician

wants to avoid failing to identify a gambler in need of treatment or referral. In this case the use of a less stringent criterion score is recommended since it will minimize the number of false negatives and thus capture most of those who are pathological gamblers. These individuals may then be followed up with more intensive testing, referral to a specialist or the implementation of treatment.

If the goal is estimating the number of pathological gamblers in the community, it makes more sense to apply a strict criterion to protect against false positive errors (e.g., Dickerson et al., 1996). In view of the unknown, but likely low, levels of help-seeking (Productivity Commission, 1999), policy makers should plan for a conservative number of pathological gamblers expected to seek treatment.

It may also be argued that when researchers present estimates of pathological gambling to policy makers in the community, they should stress interval estimates, not point estimates. Interval estimates (Gambino, 1999b) are a more reasonable measure of the accuracy of prevalence estimates (McGrath, 1998), and are recommended by the American Psychological Association in their latest guidelines for statistical reporting (Wilkinson, 1999). An interval estimate provides a measure of the degree of confidence one has that the true prevalence value has been captured by the interval. This would enable researchers to more confidently communicate their findings to funding sources and other stakeholders. It may also be noted that those stakeholders unfamiliar with the technical requirements for computing confidence intervals are, in fact, quite familiar with the concept itself. This is the result of the frequent reporting in the media of survey or poll results in which the outcome (point estimate) is presented along with an estimate of the margin of error (confidence interval), and researchers should take advantage of this equivalence to communicate the meaning of the confidence interval to stakeholders.

Setting the criterion for use by policy makers. Any decision by policy makers on the estimated number of pathological gamblers requires a rule for determining clinical or practical significance. The issue of clinical significance is a complex one (e.g., Spitzer, 1998) and the solution presented here is only one of several that may be applied. It has the advantage of being relatively simple to calculate and has a straightforward interpretation in terms of the likelihood or certainty of diagnosis. The technique is one recommended by clinical epidemiologists for making diagnostic decisions with confidence (Chu, 1999; Koch et al., 1995; Kraemer, 1992; Schmitz et al., 2000; Zhou et al., 2002). The application of this technique to the evaluation of prevalence estimates rests on the assumption that increasing score levels reflect increasing levels of severity. There is an increasing accumulation of evidence that severity is related to the likelihood that individuals will need or seek treatment (Productivity Commission, 1999). For example, in a recent national study of Australian gamblers, Tremayne et al. (2001) found that

only 12.3% of those who scored between 5 to 9 on the SOGS reported seeking help whereas 54.3% of those who scored 10 or higher sought assistance.

The method entails the calculation of the likelihood ratio (LR) where, in general terms, the LR is defined as the probability that a test result (positive or negative) would be expected in a respondent with the disorder (pathological gambler) compared to the probability that the same result would be expected in a respondent without the disorder (non pathological gambler). The LR for positive test results is therefore defined as Se / (1 – Sp), and for negative test results as (1 – Se) / Sp.

The likelihood ratio for positive (negative) tests is an empirical estimate of the power of a score or range of scores to discriminate the pathological gambler who scores positive (negative) from the non-pathological gambler who scores positive (negative). In the analysis presented below, negative predictive value is ignored as well as the LR based on negative test outcomes since with a low base-rate disorder such as pathological gambling (Shaffer et al., 1997; Welte et al., 2001), most individuals will not be pathological gamblers and these measures have little utility in this setting. A detailed discussion of the usefulness of negative predictive value and the LR for negative tests is provided by McGee (2002), Schmitz et al. (2000) and Zhou et al. (2002).

Likelihood ratios for positive tests

Computation of the LR requires a set of individuals known to have the disorder to be compared to a set of individuals known to be free of the disorder. In an ideal situation the identification of those with and those without the disorder requires the application of a gold standard (in theory, a gold standard is an errorless procedure; in practice it is that procedure considered the most accurate one available). Since gold standards do not currently exist for pathological gambling, gamblers in treatment served to define the presence of pathological gambling (sensitivity) and gamblers from the general population served to represent the absence of the disorder (1 – specificity). This is an acceptable procedure in the absence of a gold standard (Zhou et al., 2002). Although some in the general population sample may be false negatives while some gamblers in the treatment sample may be false positives, this approach is defensible since it assumes the results apply on average rather than to any specific individual (Schlesselman, 1982). There are additional problems associated with the use of the LR but these are shared with alternative methodologies. These problems have been described in more detail by Schmitz et al. (2000) and Zhou et al. (2002).

Results

The results of computing the LR are shown in Tables 2 and 3 for the SOGS and DSM-IV respectively, based on the validation data reported by Stinchfield (2002, 2003). Likelihood ratios were converted into post-test odds by use of the formula

Post-test odds = likelihood ratio (test odds) times pre-test odds, where pre-test odds = prevalence / (1 – prevalence).

Converting the result by use of post-test odds / (1 + post-test odds) results in positive predictive value. An example will be helpful. Table 2 shows that at scores of 5 or higher, Se = .985 and 1 – Sp = .017. The LR is computed as .985/.017 = 57.92 and, assuming prevalence = .019 (Welte et al., 2001), post-test odds are obtained as 57.92 times .(019/.981) = 1.1218. Positive predictive value is then obtained as 1.1218/2.1218 = .529.

The estimates of prevalence (see Tables 2–3) were obtained from the national study reported by Welte et al. (2001). A major advantage of using these estimates is that the same respondents were tested with both instruments, thus avoiding the possibility that differences in prevalence were a function of the distribution of risk factors, e.g., differences in gender, ethnicity, and co-morbidity that might occur if different samples were employed to estimate prevalence for each instrument.

The results indicate that the power to detect pathological gambling (positive predictive value) does not reach 90% until scores of nine or higher on the SOGS, and of six or higher on the DSM-IV. A recent analysis (Strong et al., 2003) using Rasch modeling (a method for obtaining equivalent measures) provides support for these results. These investigators found that scores of nine on the SOGS were equivalent to scores of six on DSM-IV.

Discussion

The results reported in Tables 2 and 3 for positive predictive value reflect, in part, the specific choice of prevalence estimates. Other researchers might select a different set of prevalence estimates and reach a different set of recommendations (Shaffer et al., 1997). The results are interpretable as indicating that it is best to employ relatively strict criteria in order to reduce or eliminate the number of false positive results, since each false positive represents an added cost to any program for which resources might be allocated. The data in Tables 2 and 3 also demonstrate that the likelihood of a diagnosis of pathological gambling increases with increasing scores, thus supporting a view that gambling lies on a continuum of severity (Shaffer & Korn, 2002). The higher the score the more likely the result will represent a true positive outcome.

A related issue bears emphasis. The results are based on two validity studies and are specific to the instruments employed by Stinchfield (2002, 2003) and to the

choice of prevalence estimates (Welte et al., 2001). This raises the important question of validity generalization (Murphy, 2003). Additional validation studies have been conducted and others are ongoing (e.g., Abbott & Volberg, 1992, 1996; Cunningham-Williams & Cottler, 2001; Fisher, 2000; Gerstein et al., 1999; Smith & Wynne, 2002; Stinchfield et al., 2001). It is unclear that the application of the LR based on other validation studies would lead to the same conclusions with respect to validating the cutoff criterion. In particular, the comparison of gamblers in treatment (on average the most severe cases) with gamblers from the general population (on average the least severe cases) is likely to result in higher–than-expected LRs than if Se and Sp were obtained from the population of interest. This is more of an issue if the use of the Se based on a clinical population is used for the purpose of estimating PPV for a non-clinical population, such as in the primary care setting (Zhou et al., 2002). It is possible to obtain measures of Se and Sp from samples from the general population, thus generalizing the procedure described here. This allows the prevalence researcher to avoid the need to conduct their own validation studies because they employed a different instrument or a variation on the instruments employed by Stinchfield (2002, 2003).

First, it must be kept in mind that validity does not refer to the test or instrument employed. Validity refers to the conclusions or inferences drawn from test scores (Rubin, 1988). The procedure described in the present analysis can be applied to any test if prevalence researchers employ independent validation criteria. Researchers routinely collect data that may serve as empirical validation criteria that are independent of the instrument employed. For example, a question that is often asked is whether or how often in the past year the gambler lost more than $100? Those who respond yes to the criterion question can be treated as equivalent to the gamblers in treatment used in the Stinchfield studies and will serve to represent true positives. Those who respond no to the question can be treated as equivalent to the general population sample used by Stinchfield and will serve to estimate false positive rates at each score level. Other questions that are also independent of the instrument can be employed, providing additional sets of LRs. It is possible to combine these by simply multiplying the respective LRs, as long as these are independent for one, two or more criteria (Sackett et al., 1991). The resulting values for PPV can be expected to be high.

The final LR may then be multiplied by pre-test odds (prevalence odds) to determine post-test odds and the resulting positive predictive value obtained. Researchers who employ a different instrument from those used by Stinchfield or a variant of these do not have to conduct their own validation study. The use of one or more independent questions provides the data required to apply the LR procedure. Once the LR has been obtained, researchers need only decide on an acceptable estimate of prevalence. These may be obtained from those available in the literature. The researcher can also compare these results from those obtained

by employing the observed sample prevalence rate. The procedure illustrated in the present analysis is therefore generalizable to other studies.

Conclusions

Policy recommendations should be based on practical (useful) and well-defined (validated) measures. Effective public health is heavily dependent on clear case-definitions that include criteria potentially categorized by the degree of certainty regarding diagnosis as “suspected” or “confirmed” (Teutsch, 1994). The present analysis indicates that when results based on scores of 10 or higher on the SOGS or six or higher on the DSM-IV are used, observers can assume with a high level of confidence that those identified as true positives are indeed pathological gamblers.

The LR and its translation into PPV is an increasingly popular methodology (Sackett et al., 1991). Researchers must develop improved measures of severity which are needed to help describe the etiology and natural history of gambling disorders (Gordis, 1996; Koeter et al., 2003; Winters et al., 1996). The data in Tables 2 and 3 demonstrate that as severity (as measured by increasing scores) increases the LR and PPV will correspondingly increase. A more relevant concern for researchers interested in policy decisions on allocation of resources is to develop better definitions of functional status and disability (Pincus et al., 1998; Spitzer, 1998). These measures may then be related to prognosis and will likely predict seeking help (Ustun & Rehm, 1998).

Current definitions should also be more strongly tied to accepted notions of clinical and social significance (Frances, 1998). Examples include: did you recently lose your job because of your gambling, does your gambling substantially interfere with important activities, how often does this occur, and what is the most recent incident? Either there are people who will benefit in terms of some non-trivial measure of quality of life if they reduce or stop their gambling or there are not. If there are, then we must decide if we wish to allocate scarce resources to help. That, in turn, requires consensus on a definition of who is a case in need of assistance?

Clearly more intensive and focused research will help to better clarify this important issue of who should be defined as a case and who should not. Future research can further refine these initial estimates, and address important issues such as taking into account the sample sizes needed to obtain sufficient power for testing hypotheses and ensuring the reliability of estimates. The question of robustness remains to be resolved. Can the present results be generalized to variants on the instruments employed here or not? In the interim, the procedures described above should serve as reasonable initial estimates.

It should be added that most errors will occur just below, at, or just above the selected cut-off value. Few researchers would argue strongly that those who score four on DSM-IV are in fact different from those who score five. Yet these individuals are generally treated differently and the researcher often behaves as though the distinction were real rather than arbitrary. All that can be known is that if the criterion is set at five then there are four possible outcomes with respect to a gambler who scores five and a gambler who scores four. The four outcomes are a) both pathological (true positive, false negative), b) neither are pathological (false positive, true negative), c) the first but not the second (true positive, true negative), or d) the second but not the first (false positive, false negative). The use of the LR or some similar procedure is applicable to any instrument, including those which may be developed prior to the adoption of a new definition for DSM-V, and thus provides a bridge between the old and the new. Those who employ current instruments and those who develop alternatives should collect and report evidence on sensitivity, specificity, as well as positive and negative predictive values, since the latter measures are much more relevant and meaningful to clinicians.

References

Abbott, M.. Volberg, R.. (1991). *Gambling and problem gambling in New Zealand. Report on phase one of the national survey*(Research Series No. 12). Wellington, New Zealand: Department of Internal Affairs.

Abbott, M.. Volberg, R.. (1992). *Frequent gamblers and problem gamblers in New Zealand*(Research Series No. 14). Wellington, New Zealand: Department of Internal Affairs.

Abbott, M. W.. Volberg, R. A.. (1996). The New Zealand National Survey of problem and pathological gambling. *Journal of Gambling Studies*, 12, 143-160.

Abbott, M. W.. Volberg, R. A.. (1999). A reply to Gambino's “An epidemiologic note on verification bias”: Implications for estimation of rates. *Journal of Gambling Studies*, 15, 233-242.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Brenner, H.. Gefeller, O.. (1997). Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Statistics in Medicine*, 16, 981-91.

Chu, K.. , (1999). An introduction to sensitivity, specificity, predictive values and likelihood ratios. *Emergency Medicine*, 11, 175-181.

Cunningham-Williams, R. M.. Cottler, L. B.. (2001). The epidemiology of pathological gambling. *Seminars in Clinical Neuropsychiatry*, 6, 155-166.

Dickerson, M.. (1993). A preliminary exploration of a two-stage methodology in the assessment of the extent and degree of gambling-related problems in the Australian population. In Eadington, W. R.. & Cornelius, J. A.. (Eds.), *Gambling behavior and problem gambling* (pp. 347-363). Reno, NV: University of Nevada, Reno.

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

Dickerson, M. G.. Volberg, R. A.. Preface/editorial for special issue. *Journal of Gambling Studies*, 12, 109-110.

Faraone, S. V.. Tsuang, M. T.. (1994). Measuring diagnostic accuracy in the absence of a “gold standard.”*American Journal of Psychiatry*, 151, 650-657.

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

Frances, A.. (1998). Problems in defining clinical significance in epidemiological studies. *Archives of General Psychiatry*, 55, 119.

Gambino, B.. (1997a). Method, method: Who's got the method? What can we KNOW about the number of compulsive gamblers? *Journal of Gambling Studies*, 13, 291-296.

Gambino, B.. (1997b). The correction for bias in prevalence estimation with screening tests. *Journal of Gambling Studies*, 13, 343-351.

Gambino, B.. (1999a). An epidemiologic note on verification bias: Implications for estimation of rates. *Journal of Gambling Studies*, 15, 223-232.

Gambino, B.. (1999b). Estimating confidence intervals and sampling proportions in two-stage prevalence designs. *Journal of Gambling Studies*, 15, 243-245.

Garrett, E. S.. Eaton, W. W.. Zeger, S.. (2002). Methods for evaluating the performance of diagnostic tests in the absence of a gold standard: A latent class model approach. *Statistics in Medicine*, 21, 1289-1307.

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

Gordis, L.. (1996). *Epidemiology*. Philadelphia: W.B. Saunders Co.

Hui, S. L.. Walter, S. D.. (1980). Estimating the error rates of diagnostic tests. *Biometrics*, 36, 67-71.

Koch, M.. Capurso, L.. Llewelyn, H.. (1995). Analyzing the discriminating power of individual symptoms, signs and test results. In Llewelyn, H.. & Hopkins, A.. (Eds.), *Analyzing how we reach clinical decisions* (pp. 51–67). London: Royal College of Physicians of London.

Koeter, M. J. W.. de Fuentes-Merillas, L.. Schippers, G. M.. van den Brink, W.. (2003). Severity of gambling addiction: Development of a new assessment instrument. *World Psychiatry*, 2, 6 (Supplement 1).

Koplan, J. P.. Thacker, S. B.. Lezin, N. A.. (1999). Epidemiology in the 21st century: Calculation, communication, and intervention. *American Journal of Public Health*, 89, 1153-1155.

Kraemer, H. C.. (1992). *Evaluating medical tests: Objective and quantitative guidelines*. Newbury Park, CA: SAGE.

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

McGee, S.. (2002). Simplifying likelihood ratios. *Journal of General Internal Medicine*, 17, 646-649.

McGrath, R. E.. (1998). Significance testing: Is there something better?*American Psychologist*, 53, 796-797.

Meehl, P. E.. Rosen, A.. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 3, 195-216.

Murphy, K. R.. (Ed.). (2003). Validity generalization: A critical review. Assessment Systems Corporation. Available at: http://www.assess.com/

National Research Council (1999). *Pathological gambling: A critical review*. Washington, DC: National Academy Press.

Pincus, H. A.. Zarin, D. A.. First, M.. “Clinical significance” and DSM IV. *Archives of General Psychiatry*, 55, 1145.

Poulin, C.. (2002). An assessment of the validity and reliability of the SOGS-RA. *Journal of Gambling Studies*, 18, 67-93.

Productivity Commission. (1999). Australia's gambling industries (Report No. 10). Canberra, Australia: AusInfo.

Rogan, W. J.. Gladen, B.. (1978). Estimating prevalence from the results of a screening test. *American Journal of Epidemiology*, 107, 71-76.

Rose, G.. Barker, D. J. P.. (1978). What is a case?*British Medical Journal*, 2, 873-874.

Rubin, D. B.. (1988). Discussion. In Wainer, H.. & Braun, H. I.. (Eds.), *Test validity* (pp. 241–256). Hillsdale, NJ: Laurence Erlbaum.

Sackett, D. L.. Haynes, R. B.. Guyatt, G. H.. Tugwell, P.. (1991). *Clinical epidemiology: A basic science for clinical medicine* (2nd ed.). Boston: Little Brown.

Schlesselman, J. J.. (1982). *Case-control studies*. New York: Oxford University Press.

Schmitz, N.. Kruse, J.. Tress, W.. (2000). Application of stratum-specific likelihood ratios in mental health screening. *Social Psychiatry & Psychiatric Epidemiology*, 35, 375-379.

Shaffer, H. J.. Hall, M. N.. Vanderbilt, J.. (1997, December 15). Estimating the prevalence of disordered gambling behavior in the United States and Canada: A meta-analysis. *Harvard Medical School Division on Addictions*. Boston: Harvard College Fellows.

Shaffer, H. J.. Korn, D. A.. (2002). Gambling and related mental disorders: A public health analysis. *Annual Review of Public Health*, 23, 171-212.

Smith, G. J.. Wynne, H. J.. (2002). Measuring gambling and problem gambling in Alberta (Final Report). Prepared for the Alberta Gaming Research Institute.

Spitzer, R. L.. (1998). Diagnosis and need for treatment are not the same. *Archives of General Psychiatry*, 55, 120.

Staquet, M.. Rozencweig, M.. Lee, Y. J.. Muggia, F. M.. (1981). Methodology for the assessment of new dichotomous diagnostic tests. *Journal of Chronic Disease*, 34, 599-610.

Stinchfield, R.. (2002). Reliability, validity, and classification accuracy of the South Oaks Gambling Screen (SOGS). *Addictive Behaviors*, 27, 1-19.

Stinchfield, R.. (2003). Reliability, validity, and classification accuracy of a measure of DSM-IV diagnostic criteria for pathological gambling. *American Journal of Psychiatry*, 160, 180-182.

Stinchfield, R.. Govoni, R.. Frisch, G. R.. (2001, November). *An evaluation of diagnostic criteria for pathological gambling*(Final Report). Windsor, ON: University of Windsor.

Strong, D. R.. Lesieur, H. R.. Breen, R. B.. Stinchfield, R.. Lejuez, C. W.. (2003). Using a Rasch model approach to examine the utility of the SOGS screen across pathological and nonpathological gamblers. *Addictive Behaviors*, 28, 1465-1472.

Teutsch, S. M.. (1994). Considerations in planning a surveillance system. In Teutsch, S. M.. & Churchill, R. E.. (Eds.), *Principles and practice of public health surveillance* (pp. 18–30). New York: Oxford University Press.

Tremayne, K.. Masterman-Smith, H.. McMillen, J.. (2001). *Survey of the nature and extent of gambling and problem gambling in the ACT*. Sydney, Australia: Australian Institute for Gambling Research, University of Western Sydney.

Ustun, T. B.. Rehm, J.. (1998). Limitations of diagnostic paradigm: It doesn't explain “need.”*Archives of General Psychiatry*, 55, 1145-1146.

Volberg, R. A.. (1994). The prevalence and demographics of pathological gamblers: Implications for public health. *American Journal of Public Health*, 84, 237-241.

Volberg, R. A.. (1999). Research methods in the epidemiology of pathological gambling: Development of the field and directions for the future. *Annuario de Psicologia*, 30, 33-46.

Walker, M. B.. Dickerson, M. G.. (1996). The prevalence of problem and pathological gambling: A critical analysis. *Journal of Gambling Studies*, 12, 233-249.

Welte, J. W.. Barnes, G. M.. Wieczorek, W. F.. Tidwell, M.. Parker, J.. (2001). Alcohol and gambling among U.S. adults: Prevalence, demographic patterns and comorbidity. *Journal of Studies on Alcohol*, 62, 706-712.

Wilkinson, L.. (1999). Statistical methods in psychology journals: Guidelines and explanations. Task Force on Statistical Inference. *American Psychologist*, 54, 594-604.

Winters, K.. Specker, S.. Stinchfield, R.. (1997). *Diagnostic interview for gambling severity (DIGS)*. Minneapolis, MN: University of Minnesota Medical School.

Zahner, G. E. P.. Hsieh, C.. Fleming, J. A.. (1995). Introduction to epidemiologic research methods. In Tsuang, M. T.. , Tohen, M.. , & Zahner, G. E. P.. (Eds.), *Textbook on psychiatric epidemiology* (pp. 23–54). New York: Wiley-Liss.

Zhou, X. H.. Obuchowski, N. A.. McClish, D. K.. (2002). *Statistical methods in diagnostic medicine*. New York: John Wiley & Sons, Inc.

Tables

Table 1

Four possible outcomes of testing

	True status		
Test results	Pathological	Not pathological	Row totals
Positive	a = TP = P*Se*N	b = FP = (1 − P)*(1 − Sp)*N	a + b
Negative	c = FN = P*(1 − Se)*N	d = TN = (1 − Sp)*Sp*N	c + d
Column totals	a + c	b + d	N

P = prevalence Se = sensitivity Sp = specificity (a + b)/N = Pp = sample prevalence estimate

Adjustment for errors may be obtained as P = [(Pp − (1 - Sp)]/ [Se − (1 - Sp)]

Number of pathological gamblers = a + c
Sensitivity = TP / (TP + FN) = a / (a + c)
Number of non-pathological gamblers = b + d
Specificity = TN / (TN + FP) = d / (b + d)
Number of positive tests = a + b
Positive Predictive Value = TP / (TP + FP) = a / (a + b)
Number of negative tests = c + d
Negative Predictive Value = TN / (TN + FN) = d / (c + d)

Table 2

Likelihood ratios (LR) and positive predictive values (PPV) based on scores on SOGS[1]

Score on SOGS	Se	1 − Sp	LR	PPV
>0	1.000	.159	6.49	.112
>1	.996	.066	15.09	.226
>2	.991	.043	23.04	.309
>3	.988	.028	35.28	.407
>4	.985	.017	57.92	.529
>5	.976	.013	75.04	.593
>6	.948	.008	118.50	.697
>7	.893	.006	148.83	.743
>8	.841	.001	841	.942
>9	.765	.000	∞	1.00

[1] Se = sensitivity, based on responses of treatment sample (N = 327); 1 − Sp = 1 − Specificity, based on responses of general population sample (N = 845); LR = Likelihood Ratio = Se / (1 − Sp); PPV = post-test odds / (1 + post-test odds) where post-test odds = pre-test odds x LR. Pre-test odds were obtained as prevalence / (1 − prevalence) employing an estimated prevalence = .019 (Welte et al, 2001).

Table 3

Likelihood ratios (LR) and positive predictive values (PPV) based on scores on DSM-IV[2]

Score level	Se	1 − Sp	LR	PPV
> 0	.992	.044	22.55	.229
> 1	.978	.024	40.75	.350
> 2	.978	.014	69.86	.480
> 3	.969	.0075	129.20	.631
> 4	.949	.0038	249.74	.767
> 5	.914	.0000	∞	1.000

[2] Se = Sensitivity, based on responses of treatment sample (N = 257); 1 − Sp = 1 − specificity, based on responses of general population sample (N = 800); LR = Likelihood Ratio = Se / (1 − Sp); PPV = post-test odds / (1 + post-test odds) where post-test odds = pre-test odds x LR. Pre-test odds were obtained as prevalence / (1 − prevalence) employing an estimated prevalence = .013 (Welte et al, 2001).

Keywords:

Keywords

prevalence estimation

,

case-definitions

,

public policy

,

validity

,

likelihood ratios

,
sensitivity
,
specificity
,
predictive value
.