

It is illegal to post this copyrighted PDF on any website.

Positive Predictive Value:

A Clinician's Guide to Avoid Misinterpreting the Results of Screening Tests

Mark Zimmerman, MD^{a,b,*}

ABSTRACT

Background: Recent studies have used self-administered screening scales in community and clinical samples to identify individuals who probably or likely had the disorder of interest. A better understanding of the statistics of screening, specifically positive predictive value, would indicate that the conclusions drawn from these studies are not justified.

Methods: The principles and statistics of diagnostic screening and how screening is distinguished from case-finding are reviewed, followed by a review of studies that have failed to consider the positive predictive value of the screening scales in the samples studied.

Results: Multiple studies of both clinical and general population samples have used screening measures as case-finding instruments. For example, two recent studies of response to electroconvulsive therapy in depressed patients used a screening scale for borderline personality disorder (BPD) and concluded that the patients with and without BPD responded equally well to treatment. However, the positive predictive value of the screening scale in these studies was less than 50%, meaning the majority of patients considered to have BPD would not have been so diagnosed if interviewed. A similar problem has also been observed in studies using screening scales for bipolar disorder in general population and primary care settings.

Conclusions: When studying a disorder with a relatively low prevalence, it is near impossible for a screening test to have sufficient positive predictive value to be used to validly compare the individuals who do and do not screen positive. Researchers using screening measures as diagnostic proxies need to discuss the issue of positive predictive value.

J Clin Psychiatry 2022;83(5):22com14513

To cite: Zimmerman M. Positive predictive value: a clinician's guide to avoid misinterpreting the results of screening tests. *J Clin Psychiatry*. 2022;83(5):22com14513.

To share: <https://doi.org/10.4088/JCP.22com14513>

© 2022 Physicians Postgraduate Press, Inc.

^aDepartment of Psychiatry and Human Behavior, Brown Medical School, Providence, Rhode Island

^bDepartment of Psychiatry, Rhode Island Hospital, Providence, Rhode Island

*Corresponding author: Mark Zimmerman, MD, 146 West River St, 11B, Providence, RI 02904 (mzimmerman@lifespan.org).

An article entitled “Positive Predictive Value” is unlikely to attract much attention from practicing clinicians. That is unfortunate, because it is as important for clinicians as for researchers to understand positive predictive value to interpret studies based on screening instruments. As will be illustrated in this article, the failure to understand the implications of the modest positive predictive value of screening instruments has resulted in inappropriate conclusions with potential public health and even treatment implications.

Let me begin with a hypothetical. A screening test for schizophrenia has been developed and it is being studied in the general population, for whom the prevalence of schizophrenia is 1%. Assume the screening test has a sensitivity of 100% and a specificity of 95%. What is the chance that a person who screens positive on the test has schizophrenia? I will answer this question later in this article, but I suggest you estimate the answer now.

A Brief Overview of the Statistics of Screening

When researchers use a screening measure and briefly summarize its prior performance, they typically refer to the scale's sensitivity and specificity. Sensitivity refers to how well the screening measure identifies individuals with the illness or disorder of interest. As illustrated in Table 1, when computing sensitivity, the numerator is the number of ill persons who are correctly identified as ill by the test, and the denominator is the total number of ill persons $[a/(a + c)]$. By contrast, specificity refers to how well the screening test identifies individuals without the illness. When computing specificity, the numerator is the number of persons without the illness who are correctly identified by the test as not having the illness, and the denominator is the total number of persons without the illness $[d/(b + d)]$.

Sensitivity and specificity indicate a screening measure's performance independent of the prevalence of the disorder in the sample. Whether the disorder is present in 5% or 50% of the sample, sensitivity and specificity are not impacted.

By contrast, positive predictive value is strongly influenced by disorder prevalence. Positive predictive value refers to the probability a person who is identified as ill by the test has the illness. When computing positive predictive value, the numerator is the same as it is in computing sensitivity (ie, the number of ill persons who are correctly identified as ill by the test); however, the denominator is now the total number of persons that the test identifies as ill $[a/(a + b)]$. Examples of the type of questions asked that refer to positive

You are prohibited from making this PDF publicly available.

Table 1. Table of Association Between a Diagnostic Gold Standard and Screening Test

Screening Test	Gold Standard Diagnosis		
	Present	Absent	Total
Positive	a	b	a + b
Negative	c	d	c + d
Total	a + c	b + d	a + b + c + d

Sensitivity = $a/(a + c)$.
 Specificity = $d/(b + d)$.
 Positive predictive value = $a/(a + b)$.
 Negative predictive value = $d/(c + d)$.

Table 2. Hypothetical Tables of Association of a Test With Sensitivity and Specificity of 90% With a Disorder Prevalence Rate of 50% or 5%

A. Disorder Prevalence = 50%

Screening Scale	Gold Standard Diagnosis		
	Present	Absent	Total
Positive	90	10	100
Negative	10	90	100
Total	100	100	200

Sensitivity = $a/(a + c) = 90/100 = 90.0\%$.
 Specificity = $d/(b + d) = 90/100 = 90.0\%$.
 Positive Predictive Value = $a/(a + b) = 90/100 = 90.0\%$.
 Negative Predictive Value = $d/(c + d) = 90/100 = 90.0\%$.

B. Disorder Prevalence = 5%

Screening Scale	Gold Standard Diagnosis		
	Present	Absent	Total
Positive	9	19	28
Negative	1	171	172
Total	10	190	200

Sensitivity = $a/(a + c) = 9/10 = 90.0\%$.
 Specificity = $d/(b + d) = 171/190 = 90.0\%$.
 Positive predictive value = $a/(a + b) = 9/28 = 32.1\%$.
 Negative predictive value = $d/(c + d) = 171/172 = 99.4\%$.

predictive value are how many women with a positive Pap smear have cervical cancer, how many men with an elevated prostate-specific antigen (PSA) level have prostate cancer, and how many individuals who are positive on a rapid screen for COVID-19 are infected with the virus. Individuals who screen positive but do not actually have the disorder are false positives.

Impact of Disorder Prevalence on Positive Predictive Value

A test's positive predictive value is higher in samples in which disorder prevalence is greater (assuming test sensitivity and specificity are fixed across samples). Consider two studies with samples of equal size but different illness prevalence rates. When test sensitivity [$a/(a + c)$] is the same, then in the sample with the higher prevalence ($a + c$), both cells a and c must be greater. Likewise, when specificity [$d/(b + d)$] remains the same, then when prevalence is higher wellness ($b + d$) correspondingly decreases and both cells b and d are lower. Positive predictive value ($a/(a + b)$) is higher in the sample with a higher prevalence because cell b is smaller.

To illustrate with some numbers, let's consider a test that has a sensitivity of 90% and specificity of 90%. In a study of

Table 3. Sample Size, Number of BPD Cases, Derived Number of True and False Positives, and Computed Positive Predictive Value of the McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD) in Two Studies of Depressed Patients' Response to Electroconvulsive Therapy

Study	Total n	No. of BPD Cases ^a	True Positive ^b	False Positive ^b	Positive Predictive Value ^b
Yip et al ¹	693	145	118	202	36.9%
Lee et al ³	137	29	24	40	37.5%

^aThe number of BPD cases is based on the MSI-BPD.

^bThe numbers of true and false positives were estimated from the value of sensitivity (81.7%) and specificity (63.2%) of the MSI-BPD based on a literature review of the scale.⁴ Positive predictive value is computed as the number of true positives divided by the sum of true positives and false positives.

Abbreviation: BPD = borderline personality disorder.

200 patients, when the prevalence of the disorder is 50%, 100 patients have the disorder ($a + c$) and 100 patients do not have the disorder ($b + d$) (Table 2A). With a sensitivity of 90%, 90 of the 100 patients with the disorder screen positive (cell a). With a specificity of 90%, 90 of the 100 patients without the disorder screen negative (cell d). If cell d equals 90, then cell b must equal 10, and the positive predictive value is 90% [$90/(90 + 10)$].

If, on the other hand, disorder prevalence is 5%, then 10 of the 200 patients have the disorder ($a + c$) and 190 do not ($b + d$) (Table 2B). With a sensitivity of 90%, 9 of the 10 patients with the disorder screen positive (cell a). With a specificity of 90%, 171 of the 190 patients without the disorder screen negative (cell d). If cell d equals 171, then cell b must equal 19, and the positive predictive value is now only 32.1% [$9/(9 + 19)$].

So, in the first instance, when the prevalence of the disorder was set at 50%, the vast majority of the patients screening positive actually had the disorder. However, when the prevalence of the disorder was low and set at 5%, then the majority of patients screening positive would *not* have the disorder.

Studies That Failed to Consider Positive Predictive Value

Two studies of response to electroconvulsive therapy. Now let's turn to some research studies that reached inappropriate conclusions because the authors failed to consider positive predictive value.

A recent study¹ reported no difference between patients with and without borderline personality disorder (BPD) in their response to electroconvulsive therapy (ECT). The authors concluded that their data "add to the evidence base that provides the clinician with a rationale for proceeding with ECT among depressed patients, notwithstanding comorbid BPD." In that study, the researchers "diagnosed" BPD with the McLean Screening Instrument for BPD (MSI-BPD).² A second recent study of the relationship between BPD and ECT³ also used the MSI-BPD to identify BPD and also found no difference in outcome between the patients who did and did not screen positive for BPD.

In both studies using the MSI-BPD, the prevalence of BPD

You are prohibited from making this PDF publicly available.

It is illegal to post this copyrighted PDF on any website.

was 21%. A review⁴ of 8 studies of the MSI-BPD in adults found that at the cutoff of 7 used by the authors of both studies, the sensitivity of the MSI-BPD was 81.7% and the specificity of the scale was 63.2%. Based on these values of sensitivity and specificity, and a prevalence of BPD of 21%, the positive predictive value of the scale is only 37% (see Table 3). That is, about two-thirds of patients in the BPD group would be false positives and would not have been diagnosed with BPD had they been interviewed with a semistructured diagnostic interview (which is the gold standard in diagnosing BPD). By contrast, in a study of ECT response in depressed patients with and without BPD that used a semistructured interview to diagnose BPD,⁵ the prevalence of BPD was 14%. Based on a BPD prevalence of 14%, and the values from the review of the studies using the MSI-BPD, the positive predictive value of the MSI-BPD would be 26.5% in these 2 studies of ECT response in depressed patients who did and did not screen positive for BPD. Thus, nearly three-quarters of the patients in the cohorts who screened positive for BPD are false positives who would not have been diagnosed with BPD if evaluated with a semistructured diagnostic interview. Neither study that used the MSI-BPD included a discussion of the limited positive predictive value of the screening test and the implication this had on any conclusions that could be drawn.

Studies of bipolar disorder. Self-report screening questionnaires have been used in general population surveys, and researchers have drawn conclusions about disorder prevalence, the psychosocial impairment and public health burden associated with the disorder, and the frequency with which disorders are underrecognized, underdiagnosed, and undertreated.

In a general population study,⁶ more than 85,000 individuals completed the Mood Disorders Questionnaire (MDQ), a self-report screening questionnaire for bipolar disorder. A subset of the subjects who screened positive and negative completed a subsequent questionnaire assessing health care utilization and medication use. The authors found that more than half of the MDQ-positive individuals were not diagnosed with bipolar disorder by treating clinicians and indicated that these clinicians “failed to detect” or “misdiagnosed” bipolar disorder in these individuals. They further examined the medications received by the MDQ-positive individuals and found that a low percentage were prescribed mood stabilizers and a higher percentage were prescribed antidepressants. The authors indicated that these patients were “inappropriately treated.” The authors begin the Discussion section of the article by stating, “The results of this US-population-based study suggest that bipolar disorder is frequently undetected or misdiagnosed, even among patients who consult psychiatrists.”^(p1532) Later in the same paragraph, the authors note, “That fewer than one in four respondents who screened positive and who consulted a primary care physician received a diagnosis of bipolar disorder is alarming...” Elsewhere in the Discussion, the authors note, “The underuse of mood stabilizers among these patients with bipolar disorder is particularly worrisome....”

The authors cite the statistics on the sensitivity and specificity of the MDQ but do not discuss the issue of limited positive predictive value. They note that, in a nonclinical sample, the MDQ had a sensitivity of 28.1% and a specificity of 97.2%. Assuming a population prevalence of bipolar disorder of 3%, the positive predictive value of the MDQ would be less than 25%. Thus, more than three-quarters of the persons they considered to have bipolar disorder because they screened positive on the scale would not have been diagnosed with bipolar disorder if interviewed.

In a study in primary care, Das et al⁷ found that nearly 10% of 1,157 patients seeking primary care at an urban general medical practice servicing low-income individuals screened positive on the MDQ. MDQ-positive patients had more depressive, anxiety, and substance use disorders; suicidal ideation; mental health care; and functional impairment. The authors noted how “remarkably” few patients had been previously diagnosed with bipolar disorder. In discussing the reasons for the low diagnostic rate, the authors offered 4 possible explanations for the low rate of previous bipolar diagnoses without mentioning low positive predictive value. Most of the Introduction and Discussion focused on the clinical and public health significance of bipolar disorder. The authors concluded that bipolar disorder is underrecognized in primary care and that primary care physicians should receive greater education about the recognition and treatment of bipolar disorder. While the authors’ overall conclusion about the underdiagnosis of bipolar disorder in primary care may well be correct, my point is that their research study cannot be used to support that conclusion because the positive predictive power of the screening instrument was low and most individuals whom the authors considered to have bipolar disorder would not have been so diagnosed if interviewed.

Humpston et al⁸ recently reported the results of the first large epidemiologic study of bipolar disorder in England. The prevalence estimate of bipolar disorder was based on the MDQ. In justifying the use of the MDQ, the authors indicated that the respective sensitivity and specificity of the MDQ in the general population were 28% and 97%. On the basis of these data, Humpston et al indicated that “individuals screening positive on the MDQ are *very likely* [italics added] to have the disorder.” That is, the sensitivity and specificity of the measure were sufficient for it to be used as a case-finding instrument in their epidemiologic study of more than 7,500 participants, though they used the term *probable bipolar disorder* in the article because follow-up diagnostic interviews were not conducted. Not only did the authors provide a prevalence estimate of bipolar disorder in England, but they also compared the individuals with probable bipolar disorder to the remainder of the sample on demographic and clinical variables. Two conclusions were drawn from their epidemiologic study.⁸ First, Humpston and colleagues indicated that the prevalence of bipolar disorder in England was similar to the rates in other parts of the world. Second, they stated that most individuals with bipolar disorder in England did not receive treatment for the disorder in the

It is illegal to post this copyrighted PDF on any website.

past year. Related to this second conclusion, the authors concluded that mental health services for bipolar disorder in England were suboptimal. The results of that study thus have potentially significant public health implications.

Humpston et al found that the prevalence of “probable bipolar disorder” (according to the MDQ) was 1.7%. Using that study’s estimate of the prevalence of bipolar disorder in England, a scale with a sensitivity of 28.1% and specificity of 97.2%, would have a positive predictive value of 14.8%. Because the vast majority of individuals who screen positive would not be diagnosed with bipolar disorder if interviewed, it was inappropriate for the authors to refer to the group that screened positive as having “probable bipolar disorder.” Moreover, with such a low positive predictive value, a comparison of patients who did and did not screen positive on the MDQ has little relevance to bipolar disorder. Several reports^{9–12} have demonstrated that individuals who screen positive on the MDQ have psychiatric diagnoses other than bipolar disorder. Thus, the authors’ conclusions regarding the suboptimal treatment of bipolar disorder in England should not be based on the data in that study. In fact, their finding that approximately 15% of the individuals who screened positive had received help for bipolar disorder closely matches the positive predictive value of the screening measure and can be interpreted as suggesting that bipolar disorder is being appropriately treated in England.

Other studies that have used the MDQ^{9,13–20} have similarly drawn conclusions about the prevalence, diagnosis, and recognition of bipolar disorder without mentioning the low positive predictive value of the measure and the likelihood that the majority of individuals who screened positive did not have bipolar disorder.

Conclusion

It is understandable why many studies use self-administered screening scales to identify the diagnosis of interest. Such scales are easy to administer, do not require interviewers to be trained to administer time-consuming evaluations, and

thus permit research to be done inexpensively. By intention, screening tests are much less costly than the more definitive diagnostic procedures. However, screening tests are designed to cast a broad net to capture most individuals who have the index disorder (ie, have high sensitivity) and are supposed to be followed by the diagnostic test to rule in or rule out the disorder. Positive Pap tests are followed by colposcopy. Elevated PSA levels are followed by prostate biopsies. For psychiatric disorders, scores above a cutoff on a screening questionnaire should be followed by a diagnostic interview. When a self-administered screening questionnaire is relied upon to identify the index diagnostic group, then the index group includes an admixture of individuals with and without the disorder of interest. When disorder prevalence is low, positive predictive value is low; thus, the majority of individuals in the index group do not actually have the disorder of interest. Clinicians need to be aware of this when reading the literature because researchers rarely discuss the positive predictive value limitation when interpreting the results of studies based on screening scales.

I will conclude by returning to the question posed at the beginning of the article. Studies posing this hypothetical question to physicians in other medical specialties have found that positive predictive value is usually overestimated.^{21,22} In a study of 10,000 subjects, a prevalence of 1% results in 100 of the 10,000 subjects having schizophrenia, and 9,900 would not be diagnosed with schizophrenia. With a sensitivity of 100%, 100 of the 100 subjects with schizophrenia would screen positive (cell a in Table 1). With a specificity of 95%, 9,405 of the 9,900 subjects who did not have schizophrenia would screen negative (cell d), and 495 subjects would screen positive (cell b). Thus, 595 subjects would screen positive (cells a + b). In this scenario, only 100 of the 595 individuals who screened positive would be diagnosed with schizophrenia. That is, the positive predictive value of a screening test with 100% sensitivity and 95% specificity in a sample in which the prevalence of the disorder was 1% would be 16.8%. Was this lower than you had predicted?

Published online: August 24, 2022.

Relevant financial relationships: None.

Funding/support: None.

REFERENCES

1. Yip AG, Ressler KJ, Rodriguez-Villa F, et al. Treatment outcomes of electroconvulsive therapy for depressed patients with and without borderline personality disorder: a retrospective cohort study. *J Clin Psychiatry*. 2021;82(2):19m13202.
2. Zanarini MC, Vujanovic AA, Parachini EA, et al. A screening measure for BPD: the McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD). *J Pers Disord*. 2003;17(6):568–573.
3. Lee JH, Kung S, Rasmussen KG, et al. Effectiveness of electroconvulsive therapy in patients with major depressive disorder and comorbid borderline personality disorder. *J ECT*. 2019;35(1):44–47.
4. Zimmerman M, Balling C. Screening for borderline personality disorder with the McLean Screening Instrument: a review and critique of the literature. *J Pers Disord*. 2021;35(2):288–298.
5. Feske U, Mulsant BH, Pilkonis PA, et al. Clinical outcome of ECT in patients with major depression and comorbid borderline personality disorder. *Am J Psychiatry*. 2004;161(11):2073–2080.
6. Frye MA, Calabrese JR, Reed ML, et al. Use of health care services among persons who screen positive for bipolar disorder. *Psychiatr Serv*. 2005;56(12):1529–1533.
7. Das AK, Olfson M, Gameraff MJ, et al. Screening for bipolar disorder in a primary care practice. *JAMA*. 2005;293(8):956–963.
8. Humpston CS, Bebbington P, Marwaha S. Bipolar disorder: prevalence, help-seeking and use of mental health care in England. Findings from the 2014 Adult Psychiatric Morbidity Survey. *J Affect Disord*. 2021;282:426–433.
9. Amin-Esmaeili M, Motevalian A, Rahimi-Movaghar A, et al. Bipolar features in major depressive disorder: results from the Iranian Mental Health Survey (IranMHS). *J Affect Disord*. 2018;241:319–324.
10. Parker G, Graham R, Rees AM, et al. A diagnostic profile of those who return a false positive assignment on bipolar screening measures. *J Affect Disord*. 2012;141(1):34–39.
11. Zimmerman M, Galione JN, Ruggero CJ, et al. Screening for bipolar disorder and finding borderline personality disorder. *J Clin Psychiatry*. 2010;71(9):1212–1217.
12. Zimmerman M, Chelminski I, Dalrymple K, et al. Screening for bipolar disorder and finding borderline personality disorder: a replication and extension. *J Pers Disord*. 2019;33(4):533–543.
13. Olfson M, Das AK, Gameraff MJ, et al. Bipolar depression in a low-income primary care clinic. *Am J Psychiatry*. 2005;162(11):2146–2151.
14. Carta MG, Cardia C, Mannu F, et al. The high

It is illegal to post this copyrighted PDF on any website.

- frequency of manic symptoms in fibromyalgia does influence the choice of treatment? *Clin Pract Epidemiol Ment Health*. 2006;2(1):36.
15. Wilke WS, Gota CE, Muzina DJ. Fibromyalgia and bipolar disorder: a potential problem? *Bipolar Disord*. 2010;12(5):514–520.
 16. Lim C, Olson J, Zaman A, et al. Prevalence and impact of manic traits in depressed patients initiating interferon therapy for chronic hepatitis C infection. *J Clin Gastroenterol*. 2010;44(7):e141–e146.
 17. Kiejna A, Pawlowski T, Dudek D, et al. The utility of Mood Disorder Questionnaire for the detection of bipolar diathesis in treatment-resistant depression. *J Affect Disord*. 2010;124(3):270–274.
 18. Dudek D, Rybakowski JK, Siwek M, et al. Risk factors of treatment resistance in major depression: association with bipolarity. *J Affect Disord*. 2010;126(1-2):268–271.
 19. Masters GA, Brenckle L, Sankaran P, et al. Positive screening rates for bipolar disorder in pregnant and postpartum women and associated risk factors. *Gen Hosp Psychiatry*. 2019;61:53–59.
 20. Ghanbari Jolfaei A, Ataei S, Ghayoomi R, et al. High frequency of bipolar disorder comorbidity in medical inpatients. *Iran J Psychiatry*. 2019;14(1):60–66.
 21. Manrai AK, Bhatia G, Strymish J, et al. Medicine's uncomfortable relationship with math: calculating positive predictive value. *JAMA Intern Med*. 2014;174(6):991–993.
 22. Morgan DJ, Pineles L, Owczarzak J, et al. Accuracy of practitioner estimates of probability of diagnosis before and after testing. *JAMA Intern Med*. 2021;181(6):747–755.

You are prohibited from making this PDF publicly available.