

Published in final edited form as:

Neuroepidemiology. 2013 ; 41(0): 208–216. doi:10.1159/000354629.

Is it time to rethink how neuropsychological tests are used to diagnose mild forms of HIV-associated neurocognitive disorders? Impact of false-positive rates on prevalence and power

Ana-Claire L. Meyer, MD^{1,2}, W. John Boscardin, PhD³, Judith K. Kwasa, MMed², and Richard W. Price, MD¹

¹Department of Neurology, San Francisco General Hospital, University of California, San Francisco, San Francisco, CA, USA

²Center for Microbiology Research, Kenya Medical Research Institute, Nairobi, Kenya

³Division of Geriatrics and Division of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA

Abstract

Background—Between 0–48% of normal HIV-uninfected individuals score below threshold neuropsychological test scores for HIV-associated neurocognitive disorders (HAND), or are false-positives. There has been little effort to understand the effect of varied interpretations of research criteria for HAND on false-positive frequencies, prevalence and analytic estimates.

Methods—The proportion of normal individuals scoring below Z-score thresholds drawn from research criteria for HAND, or false-positive frequencies, was estimated in a normal Kenyan population and a simulated normal population using varied interpretations of research criteria for HAND. We calculated the impact of false-positive frequencies on prevalence estimates and statistical power.

Corresponding Author: Ana-Claire L. Meyer, U.S.: Department of Neurology, University of California, San Francisco, San Francisco General Hospital, 1001 Potrero Avenue, Building 1, Room 101, San Francisco, CA 94110. off: +1-415-206-4449, fax: +1-415-206-4055, meyerac@sfg.ucsf.edu, Kenya: Family AIDS Care and Education Services, KEMRI, Box 614-40100, Kisumu KENYA, tel: +254-715-059-933.

W. John Boscardin, University of California San Francisco, 4150 Clement Street, San Francisco, CA 94143, john.boscardin@ucsf.edu
Judith K. Kwasa, University of Nairobi, Box 52535-00200 City Square, Nairobi, Kenya, jkwasa@yahoo.com

Richard W. Price, Department of Neurology, University of California, San Francisco, San Francisco General Hospital, 1001 Potrero Avenue, Building 1, Room 101, San Francisco, CA 94110, rwprice@sfg.ucsf.edu

Institute where work was conducted: Family AIDS Care and Education Services Research, Care and Training Program Centre for Microbiology Research Kenya Medical Research Institute Kisumu, Kenya

Conflict of Interest

Ana-Claire Meyer has no conflicts of interest to report.

John Boscardin has no conflicts of interest to report.

Judith Kwasa has no conflicts of interest to report.

Richard Price has no conflicts of interest to report.

Author Contributions

Ana-Claire Meyer designed the study, collected the data, conducted the data analysis, interpreted the data, and wrote the manuscript.

John Boscardin conducted the simulations, provided statistical consultation and critically reviewed the manuscript.

Judith Kwasa collected the data, conducted the data analysis, and critically reviewed the manuscript.

Richard Price designed the study, interpreted the data and critically reviewed the manuscript.

Results—False-positive frequencies of 2–74% were observed for Asymptomatic Neurocognitive Impairment/Mild Neurocognitive Disorder and 0–8% for HIV-associated Dementia. False-positive frequencies depended on definition of an abnormal cognitive domain, Z-score thresholds, and neuropsychological battery size. Misclassification led to clinically important overestimation of prevalence and dramatic decreases in power.

Conclusions—Minimizing false-positive frequencies is critical to decrease bias in prevalence estimates and minimize reductions in power in studies of association, particularly for mild forms of HAND. We recommend changing the Z-score threshold to -1.5 for mild impairment, limiting analysis to 3–5 cognitive domains, and using the average Z-score to define an abnormal domain.

Keywords

Africa; HIV; dementia; prevalence; power

Introduction

Considerable debate surrounds the diagnosis of mild forms HIV-associated neurocognitive disorders (HAND).^{1, 2} Direct validation of the criteria for Asymptomatic Neurocognitive Impairment (ANI) and Mild Neurocognitive Disorder (MND), which rely largely or exclusively on neuropsychological testing, has been challenging (Appendix 1).³ There are no adequately powered longitudinal clinical-pathological correlation studies^{4, 5} and there is no gold standard antemortem biomarker or imaging finding.

The prevalence of ANI/MND ranges from 26–76% and HAD from 1–35%. Studies of neuropsychological batteries used for HAND in normal HIV-uninfected populations suggest this heterogeneity may be due, at least in part, to measurement error.^{6–12} For example, 15–22% of an HIV-uninfected control group⁶ and 20% of a simulated normal population⁷ had neuropsychological test scores below the threshold for HAND, which we call false-positive cases. In addition, 4–23% of a normal population had Z-scores -1 and 0–6% had Z-scores -2 on individual tests (unpublished work⁸). While it has been argued that a 15% false-positive frequency is acceptable for neuropsychological tests,⁹ the impact of this measurement error on prevalence and analytic estimates for HAND has not been explored quantitatively.

Furthermore, neuropsychological batteries will have higher false-positive frequencies than individual tests because they involve multiple comparisons. Multiple comparisons occur when one compares two groups using multiple outcomes and attention is paid to the “strongest” differences. Here, the more comparisons made, the more likely one is to find a “strong difference,” or statistically significant outcome.¹⁰ With neuropsychological batteries, when a battery of many tests is administered and scores are compared to a normal population (multiple outcomes), and diagnoses are based on the most abnormal Z-scores (the strongest differences), the more tests that are done, the more likely it is to find two abnormal Z-scores (diagnose a normal individual as impaired).

False-positive cases will lead to biased prevalence estimates and reductions in power for analytical estimates. A false-positive on a diagnostic test, here a battery of

neuropsychological tests, can lead to a form of measurement error called non-differential misclassification.¹¹ Non-differential misclassification arises from errors in classification which occur across all levels of the variable in question. For example, in a randomized clinical trial of a new medication for HAND, the treatment and control groups would have an equal probability of having false-positive cases assuming the only contributor to false-positive cases was the neuropsychological tests. In general, non-differential misclassification will bias tests of association to the null.^{11, 12} Misclassification does not affect the validity of statistical tests (Type I error), but can drastically reduce their power (Type II error).¹³ Using the example above, non-differential misclassification will not lead investigators to falsely conclude a new treatment for HAND is effective, but it could lead investigators to falsely conclude a new treatment is not effective, when in truth it is effective. Thus, misclassification, or false-positive diagnoses of ANI/MND, could potentially affect a broad range of research including studies of prevalence, biomarkers, functional imaging, risk factors, or the effect of treatments.

Thus, in this study it is our goal to: (1) estimate misclassification (the false-positive frequency) on a neuropsychological battery for HAND using empiric and theoretical methods; (2) explore the impact of multiple comparisons (neuropsychological battery size) on the false-positive frequency; (3) explore the impact of the false-positive frequency on prevalence estimates; and (4) explore the impact of the false-positive frequency on power in studies of association.

Methods

Misclassification: Empirical estimation of the false-positive frequency

Between October 2009 and March 2010, we enrolled 84 HIV-uninfected Kenyan adults without neurologic comorbidities, depression, substance abuse, or a suspected learning disorder. Details of the population and methods can be found in our prior work (unpublished⁸). The neuropsychological test battery covered seven cognitive domains recommended by the Frascati criteria³ and contained a total of seventeen tests which were selected to parallel the neuropsychological battery used in the CHARTER study.¹⁴ We determined the proportion of normal individuals who met neuropsychological criteria for HAND using Frascati criteria (Appendix 1):³ (1) ANI/MND (two domains 1 standard deviation (SD) below mean) and (2) HAD (two domains 2SD below mean). Because Frascati criteria do not define an “abnormal domain,” we explored four different definitions of an abnormal domain: (1) one abnormal test in a domain; (2) two abnormal tests in a domain; (3) all tests in a domain must be abnormal; (4) the average of all test scores across a domain was abnormal. Z-scores were truncated at ± 3 for the purpose of calculating averages to avoid undue influence of a single extreme test score.

Misclassification: Theoretical estimation of the false-positive frequency

The false-positive frequency for a single neuropsychological test in a normal population is based largely on the statistical properties of the norming process (Figure 1).⁹ Raw scores on neuropsychological tests are converted to Z-scores based on representative population data.^{15, 16} The distribution of Z-scores should theoretically be normal.^{15, 16} Thus, 15.9% of a

normal population should have Z-scores -1 , 6.7% should have Z-scores -1.5 , and 2.2% should have Z-scores -2 . If neuropsychological test scores are independent of one another, simple multiplication could be used to calculate the expected false-positive frequency for a battery--akin to calculating the probability of two heads in a coin toss. However, neuropsychological test scores are correlated both within and between cognitive domains which makes such calculations more complex.

Thus, we simulated a normal population using the structure of the neuropsychological test battery described above (7 cognitive domains, 17 individual test scores). We assumed correlated normal distributions around a true average score of zero. Z-scores were truncated at ± 3 for the purpose of calculating averages. We explored a range of correlation structures both within and between domains and reported results for low correlation (0 between; 0 within); moderate correlation (0.5 between; 0.5 within); and high correlation (0.9 between; 0.9 within). We also analyzed results using an empiric correlation structure derived from the actual between- and within-domain correlation derived from our Kenyan sample described above. We estimated the proportion of normal individuals who would meet criteria for HAND using the previously mentioned definitions of an abnormal domain.

Multiple Comparisons: The effect of neuropsychological battery size on the false-positive frequency

We also explored the impact of different numbers of domains and tests per domain on the false-positive frequency. We varied the number of domains from two to ten, and the number of tests per domain from two to five. To minimize the effect of test number, we used the average scores across a domain to define an abnormal domain. We performed simulations using two correlation structures.. The first was derived from the average correlation within and between domains in the Kenyan sample (0.31 within; 0.17 between). The second was the moderate correlation structure that was chosen as its performance was similar to that observed in the Kenyan sample (within=0.5, between=0.5).

Simulations were performed in R (v.2.13.2, Vienna, Austria) using 100,000 repetitions resulting in Monte Carlo error of ± 0.2 –0.4% for the estimated proportions.

Exploring the impact of the false-positive frequency on prevalence and power

This analysis was based on a classic paper by Bross¹³ which has had wide influence over time.^{11, 17} Bross derives formulas for the direct calculation of: (1) bias in prevalence estimates, and (2) effective sample size (Appendix 2). Reduction in power can be expressed practically as a reduction in effective sample size.¹³ For example, in a study of 200 individuals, the loss in power due to misclassification might be equivalent to a “loss” of 20 individuals, leaving an effective sample size of 180. The effective sample size is a function of the false-positive frequency, false-negative frequency, the true prevalence of disease, and the actual sample size. We investigated a range of estimates relevant to the diagnoses of HAND. We used prevalence rates of 2%, 10%, 25%, and 40% chosen to be representative of HAD, MND, and the latter two of ANI.^{14, 18} We used false-positive frequencies in the ranges observed in our prior analyses (1–40%). We assumed a false-negative frequency of

5%. Sensitivity analyses were performed with a false-negative frequency of 25% (data not shown).

Ethics Statement—The protocol was approved by the University of California San Francisco Committee on Human Research and the National Ethical Review Committee of the Kenya Medical Research Institute. Written informed consent was obtained from each participant.

Role of the Funding Source—The funders did not have any role in study design, collection, analysis or interpretation of data, writing the report, or decision to publish.

Results

Misclassification: Empirical estimation of the false-positive frequency

Between 8–48% of normal Kenyan adults met criteria for ANI/MND (Z-scores -1 SD in 2 domains) (Figure 2; Appendix 3-Table 1). The highest false-positive frequencies were obtained when one abnormal test score defined an abnormal domain. The lowest false-positive frequencies were obtained when all the test scores in a domain were required to be abnormal. False-positive frequencies obtained using the average score or two abnormal tests per domain were in the middle range. A more stringent definition for mild impairment (Z-scores -1.5 in 2 domains) resulted in lower false-positive frequencies. With the exception of the most permissive criteria (one abnormal test score), no normal individuals met criteria for HAD (Z-scores -2 in 2 domains).

Misclassification: Theoretical estimation of the false-positive frequency

False-positive frequencies obtained using a simulated normal population were similar to the Kenyan normal adults described previously, particularly those using moderate correlation or the empirical correlation structure derived from the Kenyan population (Figure 2; Appendix 3-Table 2). Depending on the correlation structure and the definition of an abnormal domain, false-positive frequencies for ANI/MND (Z-scores -1) ranged from 2–74%. Again, the highest false-positive frequencies were observed when one abnormal test defined an abnormal domain, and the lowest when requiring all tests to be abnormal. False-positive frequencies for average scores and two abnormal tests were again in the middle range. Using more stringent criteria for mild impairment (Z-scores -1.5), false-positive frequencies ranged from 0–29%, and using criteria for HAD (Z-scores -2), false-positive frequencies ranged from 0–8%.

Multiple Comparisons: The effect of neuropsychological battery size on the false-positive frequency

As expected, the larger the number of domains, the higher the false-positive frequency for each of the Z-score thresholds (Figure 3; Appendix 3-Table 3). When averaging the scores across a domain, the number of tests per domain had only small effects on the false-positive frequency. As with prior analyses, the highest false-positive frequencies were observed for ANI/MND (Z-scores -1) and ranged from 4–41%. Stringent criteria for mild impairment

(Z-scores -1.5) resulted in false-positive frequencies of 1–16%, and criteria for HAD (Z-scores -2) in false-positive frequencies of 0–5%.

Exploring the impact of the false-positive frequency on prevalence

Higher false-positive frequencies led to greater bias in measured prevalence, or greater overestimation of prevalence (Figure 4; Appendix 4-Table 1). For example, a false-positive frequency of 1% with a true prevalence of disease of 2% (like the prevalence of HAD), would lead to an estimated or observed prevalence frequency of 3%, or a bias of 1%. The differences grew with higher false-positive frequencies such that a false-positive frequency of 20% with a true prevalence of 2% led to measured prevalence of 22% (bias of 20%). However, the magnitude of the bias was attenuated with higher true prevalence. Thus, when the true prevalence was 10% (like MND), the measured prevalence was 28% (bias of 18%). Similarly, with a true prevalence of 25% (like ANI), the measured prevalence was 39% (bias of 14%).

Exploring the impact of the false-positive frequency on power

Even low false-positive frequencies had striking effects on power when the true prevalence of disease was low (2%) as for HAD (Figure 5; Appendix 4-Table 2). Here, a false-positive frequency as low as 1% had large effects on effective sample size—reducing an actual sample size of 100 to an effective sample size of 55. This reduction was magnified for larger sample sizes—an actual sample size of 1000 was reduced to 186. Similarly dramatic effects were observed with false-positive frequencies typical of those observed for ANI/MND. A false-positive frequency of 20% and true prevalence of 10% (like MND), reduced an actual sample size of 100 to 3 and from 1000 to 10. The same false-positive frequency of 20% and a higher true prevalence of 25% (like ANI), led to comparable reductions from 100 to 9 and from 1000 to 10. Finally, a false-positive frequency of 40% and a true prevalence of 25% (like the highest estimate for ANI) reduced sample sizes from 100 to 2 and 1000 to 2.

Discussion

Current research criteria for the diagnosis of ANI/MND result in high false-positive frequencies but not for HAD. The high false-positive frequencies are due to a combination of non-differential misclassification and multiple comparisons which are intrinsic properties of neuropsychological batteries and the scoring process. High false-positive frequencies lead to unacceptably high bias in prevalence estimates and a dramatic loss in power. Thus, it is critical to minimize the false-positive frequency for neuropsychological batteries. We propose striving for a false-positive frequency of less than 5% for ANI/MND, leading to moderate reductions in power at a true prevalence of 25% (from 100 to 72 and 1000 to 227). Further we propose striving for a false-positive frequency of less than 1% for HAD to minimize the reductions in power. In the paragraphs that follow, we make specific recommendations about the structure of neuropsychological batteries and threshold scores that will achieve these false-positive frequencies.

Minimizing Misclassification

High false-positive frequencies were observed in actual and simulated normal populations, and varied depending on the definition of an abnormal cognitive domain. In addition, false-positive frequencies using a threshold Z-scores -1 were unacceptably high regardless of the parameters used. Thus we recommend changing the threshold Z-score for ANI/MND to -1.5 . A threshold Z-score of -2 is acceptable for HAD. Further, we generally recommend using a definition for an abnormal cognitive domain that minimizes the false-positive frequency, though other considerations for the optimal definition are discussed in sections that follow.

Minimizing Multiple Comparisons

In general, more tests per domain and a greater number of domains result in higher false-positive frequencies. Thus, to minimize the effect of test number on the false-positive frequency, we recommend defining an abnormal domain as the average Z-score across that domain, even though this definition did not lead to the lowest false-positive frequencies we observed. This would also enable flexibility in designing neuropsychological batteries for different purposes and help comparability across studies. In addition, we recommend grouping individual tests into no more than three to five cognitive domains.

Parallels with Mild Cognitive Impairment (MCI) due to Alzheimer's Disease

Like HAND, the diagnosis of Mild Cognitive Impairment (MCI) due to early Alzheimer's Disease rests primarily on neuropsychological testing with a threshold Z-score between -1 and -1.5 ^{19, 20} and only a few small clinical-pathologic correlation studies exist.²¹ Depending on the MCI definition used, there is substantial variation in the prevalence from 4%–42%.¹⁸ There are high false-positive frequencies for MCI; nearly half of healthy older adults will have borderline test scores ($-1.3SD$)²² and 0–26% will meet criteria for possible or probable memory impairment.²³ Furthermore, in longitudinal studies of MCI, the positive predictive value for Alzheimer's dementia varies from 5–48%, and frequencies of reversion to normal cognition range from 9–67%.¹⁸ While the empiric data used for this analysis focused on HAND, the results of the simulations are largely independent of the demographic characteristics of the normal population or the tests comprising the neuropsychological battery. Thus, the results of this study are also likely applicable to the diagnostic criteria for MCI. It is critical to recognize that both MCI and ANI/MND are clinical classifications which may or may not be the phenotypic manifestations of the presumed underlying pathology, Alzheimer's Disease and HIV-associated dementia respectively. Ultimately, identifying biomarkers and/or imaging findings will likely be necessary components to the definitive diagnosis of these disorders.

Limitations

The first major limitation of this study is that the empiric data is derived from one small population in Western Kenya. However, our estimates of false-positive frequencies are comparable to those from much larger studies⁶ as well as the simulations presented. The second major limitation is that this study focused only on the false-positive frequency due to criteria for neuropsychological testing. Although the definition of ANI is based exclusively

on neuropsychological testing, the definition of MND and HAD additionally require symptoms of cognitive impairment and/or impairment in function.²⁰ Thus, the false-positive frequencies obtained using the full criteria will likely be lower than those obtained from neuropsychological testing alone. Furthermore, this analysis does not address false positive diagnoses from ‘confounding conditions’, such as substance abuse or depression, which can impair test performance but are unrelated to the direct effects of HIV infection.¹⁴

While a neuropsychological battery that is both highly sensitive and specific would be ideal, this is difficult to achieve. When we arbitrarily define a cut-off point in our test (i.e. choose a Z-score), specificity (low frequency of false-positives) can only be improved at the expense of sensitivity.²⁴ As described previously, we cannot directly calculate the sensitivity of the neuropsychological batteries for HAND, because there is no antemortem gold standard biomarker or imaging finding. Thus, lowering the Z-score threshold to -1.5 to improve specificity may decrease sensitivity though it is difficult to estimate by how much. In addition, it will be important to explore in detail the statistical properties of other summary measures of neuropsychological performance such as the Global Deficit Score²⁵ or the average Z-score, and the effects of misclassification on other epidemiological parameters such as incidence.

Implications for current research

This analysis demonstrates that using the Frascati definition of ANI/MND results in a high frequency of false-positive diagnoses and that these false-positive diagnoses lead us to overestimate the prevalence of these disorders. Furthermore, this analysis found that including individuals with the Frascati definition of ANI/MND in studies of association dramatically decreases power. This decrease in power is applicable across study types; studies of biomarkers, functional imaging, risk factors, or treatment effects will all be affected. Ultimately, this decrease in power will lead investigators to falsely conclude there is no association between HAND and a particular biomarker, imaging finding, risk factor or treatment when in fact there is a true association. Thus, there is an urgent need both to revisit the Frascati criteria and possibly re-analyze studies which include individuals with ANI/MND as part of HAND.

Conclusion

Current research criteria for the diagnosis of ANI/MND have high false-positive frequencies, while false-positive frequencies for HAD are low. False-positive frequencies are a result of Z-score thresholds, neuropsychological battery size (number of tests and domains), and the definition of an abnormal cognitive domain. False-positive frequencies in the ranges we have described have dramatic impacts on statistical power and prevalence estimates for ANI/MND as well as important impacts on statistical power for HAD.

Minimizing false-positive frequencies is critical to decrease bias in prevalence estimates and decrease Type II error in studies of association. We recommend: (1) changing the Z-score threshold to -1.5 for ANI/MND (2) limiting analysis of neuropsychological batteries to three to five cognitive domains (3) using average Z-score to define an abnormal domain.

Additional research to determine the impact of these changes on the sensitivity of these criteria is essential. Finally, it is critical to further explore the relevance of these findings to disorders like MCI which use similar diagnostic criteria.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Jeff Martin, University of California San Francisco, for his critical review and comments on this manuscript. This study was supported by the American Academy of Neurology Practice Research Training Fellowship (Meyer), the Fogarty International Center of the National Institutes of Health (K01TW008764), the Fogarty International Clinical Research Fellowship (5 R24 TW00798; 3 R24 TW00798-02S1) (Meyer, Kwasa) from the National Institutes of Health, Fogarty International Center through Vanderbilt University, the National Cancer Institute (NCI), the National Institute on Drug Abuse (NIDA), Office of the Director, National Institutes of Health and the National Institute of Mental Health (NIMH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This paper is published with the permission of the Director of KEMRI.

References

1. McArthur JC, Brew B. HIV-associated neurocognitive disorders: is there a hidden epidemic? *AIDS*. 2010; 24:1367–1370. [PubMed: 20559041]
2. Torti C, Foca E, Cesana B, Lescure F. Asymptomatic neurocognitive disorders in patients infected with HIV: fact or fiction. *BMC Medicine*. 2011; 9:138. [PubMed: 22204606]
3. Antinori A, Arendt G, Becker J, et al. Updated research nosology for HIV-associated neurocognitive disorders. *Neurology*. 2007; 69:1789–1799. [PubMed: 17914061]
4. Cherner M, Cysique L, Heaton RK, et al. Neuropathologic confirmation of definitional criteria for human immunodeficiency virus-associated neurocognitive disorders. *Journal of Neurovirology*. 2007; 13:23–28. [PubMed: 17454445]
5. Cherner M, Masliah E, Ellis RJ, et al. Neurocognitive dysfunction predicts postmortem findings of HIV encephalitis. *Neurology*. 2002; 59:1563–1567. [PubMed: 12451198]
6. Cysique L, Franklin D, Abramson I, et al. Normative data and validation of a regression based summary score for assessing meaningful neuropsychological change. *Journal of Clinical and Experimental Neuropsychology*. 2011; 33(5):505–522. [PubMed: 21391011]
7. Gisslen M, Price RW, Nilsson S. The definition of HIV-associated neurocognitive disorders: are we overestimating the prevalence? *BMC Infectious Diseases*. 2011; 11:356. [PubMed: 22204557]
8. Cettomai D, Boscardin W, Kwasa J, et al. A Cost-Effective Method to Develop Neuropsychological Test Norms for a Battery Designed for the Diagnosis of HIV-Associated Neurocognitive Disorders. 2012
9. Heaton, R.; Chelune, G.; Talley, J.; Kay, G.; Curtiss, G. Revised Comprehensive Norms for an Expanded Halstead-Reitan Battery: Demographically Adjusted Neuropsychological Norms for African American and Caucasian Adults. Lutz, FL: Psychological Assessment Resources, Inc.; 2004.
10. Sainani K. The problem of multiple testing. *PM&R*. 2009; 1:1098–1103. [PubMed: 20006317]
11. Hofler M. The effect of misclassification on the estimation of association: a review. *International Journal of Methods in Psychiatric Research*. 2005; 14(2):92–101. [PubMed: 16175878]
12. Jurek A, Greenland S, Maldonado G, Church T. Proper interpretation of non-differential misclassification effects: expectation vs observations. *International Journal of Epidemiology*. 2005; 34:680–687. [PubMed: 15802377]
13. Bross I. Misclassification in 2×2 tables. *International Biometric Society*. 1954; 10(4):478–486.

14. Heaton RK, Clifford DB, Franklin DR, et al. HIV-associated neurocognitive disorders persist in the era of potent antiretroviral therapy: CHARTER Study. *Neurology*. 2010 Dec 7; 75(23):2087–2096. [PubMed: 21135382]
15. Spreen, O.; Strauss, E. *Compendium of Neuropsychological Tests*. New York: Oxford University Press; 1998.
16. Strauss, E.; Sherman, E.; Spreen, O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. Third ed.. New York: Oxford University Press Inc.; 2006.
17. Flegal K, Brownie C, Haas J. The effects of exposure misclassification on estimates of relative risk. *American Journal of Epidemiology*. 1986; 123(4):736–751. [PubMed: 3953551]
18. Ward A, Arrighi H, Michels S, Cedarbaum J. Mild cognitive impairment: Disparity of incidence and prevalence estimates. *Alzheimers & Dementia*. 2012; 8:14–21.
19. Albert M, DeKosky S, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute of Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011; 7(3):270–279. [PubMed: 21514249]
20. Mayeux R, Reitz C, Brickman A, et al. Operationalizing diagnostic criteria for Alzheimer's disease and other age-related cognitive impairment-Part 1. *Alzheimers & Dementia*. 2011; 7:15–34.
21. Sabbagh M, Cooper K, DeLange J, et al. Functional, Global and Cognitive Decline Correlates to Accumulation of Alzheimer's Pathology in MCI and AD. *Curr Alzheimer Res*. 2010; 7(4):280–286. [PubMed: 19715548]
22. Palmer B, Boone K, Lesser I, Wohl M. Base Rates of "Impaired" Neuropsychological Test Performance Among Healthy Older Adults. *Arch Clin Neuropsychol*. 1988; 13(6):503–511. [PubMed: 14590634]
23. Brooks B, Iverson G, Feldman H, Holdnack J. Minimizing Misdiagnosis: Psychometric Criteria for Possible or Probably Memory Impairment. *Dement Geriatr Cogn Disord*. 2009; 27:439–450. [PubMed: 19401611]
24. Fletcher, R.; Fletcher, S. *Clinical Epidemiology: The Essentials*. 4th edition. Baltimore: Lippincott, Williams & Wilkins; 2005. p. 40-43.
25. Carey C, Woods S, Gonzalez R, et al. Predictive Validity of Global Deficit Scores in Detecting Neuropsychological Impairment in HIV Infection. *Journal of Clinical and Experimental Neuropsychology*. 2004; 26(3):307–319. [PubMed: 15512922]

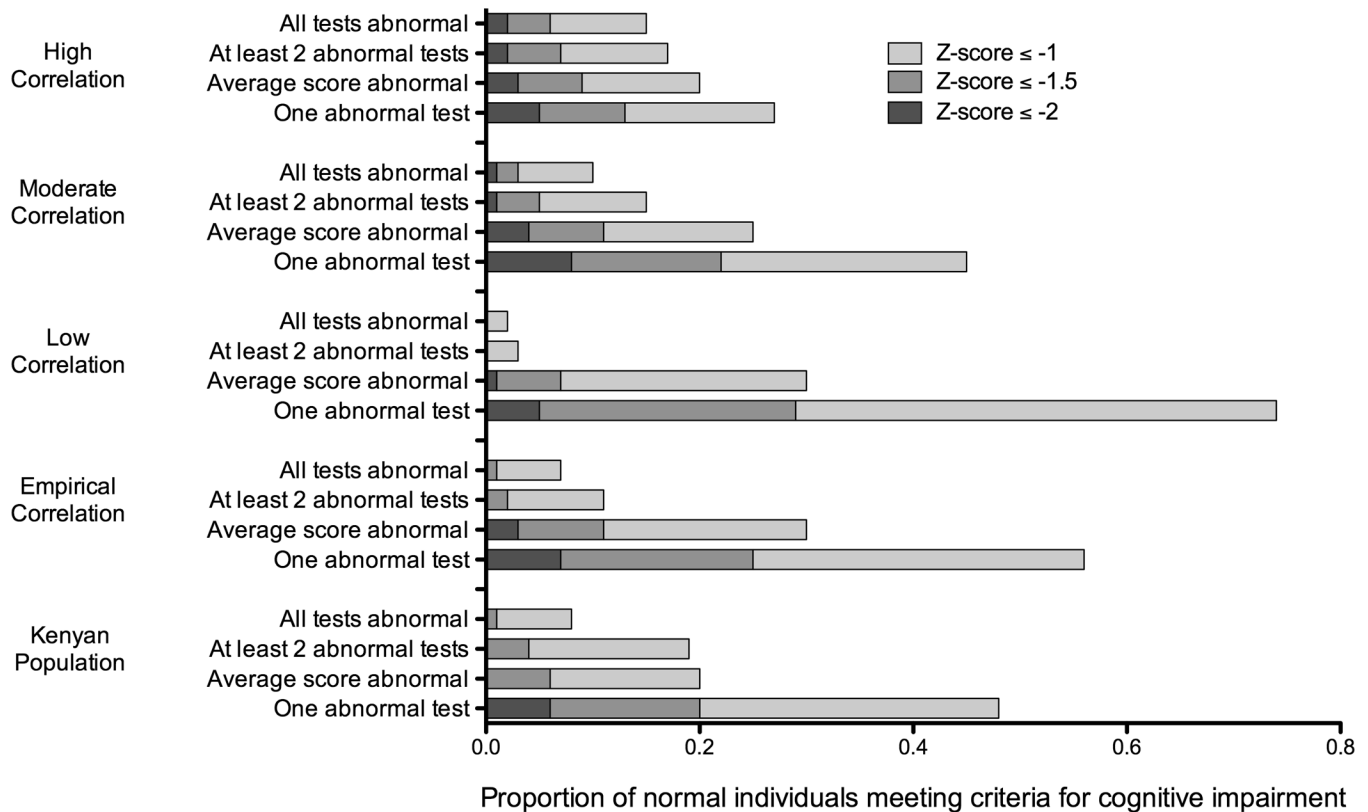


Figure 1.

The proportion of a Kenyan normal adults and a simulated normal population* who meet criteria for cognitive impairment using various definitions of impairment and various criteria for an abnormal domain.

* Simulations were based on a neuropsychological test battery identical to the one used for the Kenyan population which tested seven cognitive domains with seventeen individual tests. Z-scores were truncated at ± 3 . Simulations assumed correlated normal distributions around a true average score ($SD=0$). A range of correlation structures were explored including an empirical correlation structure derived from the actual correlations observed in the Kenyan sample. Simulations were performed using 100,000 repetitions leading to a standard error of approximately ± 0.2 – 0.4% .

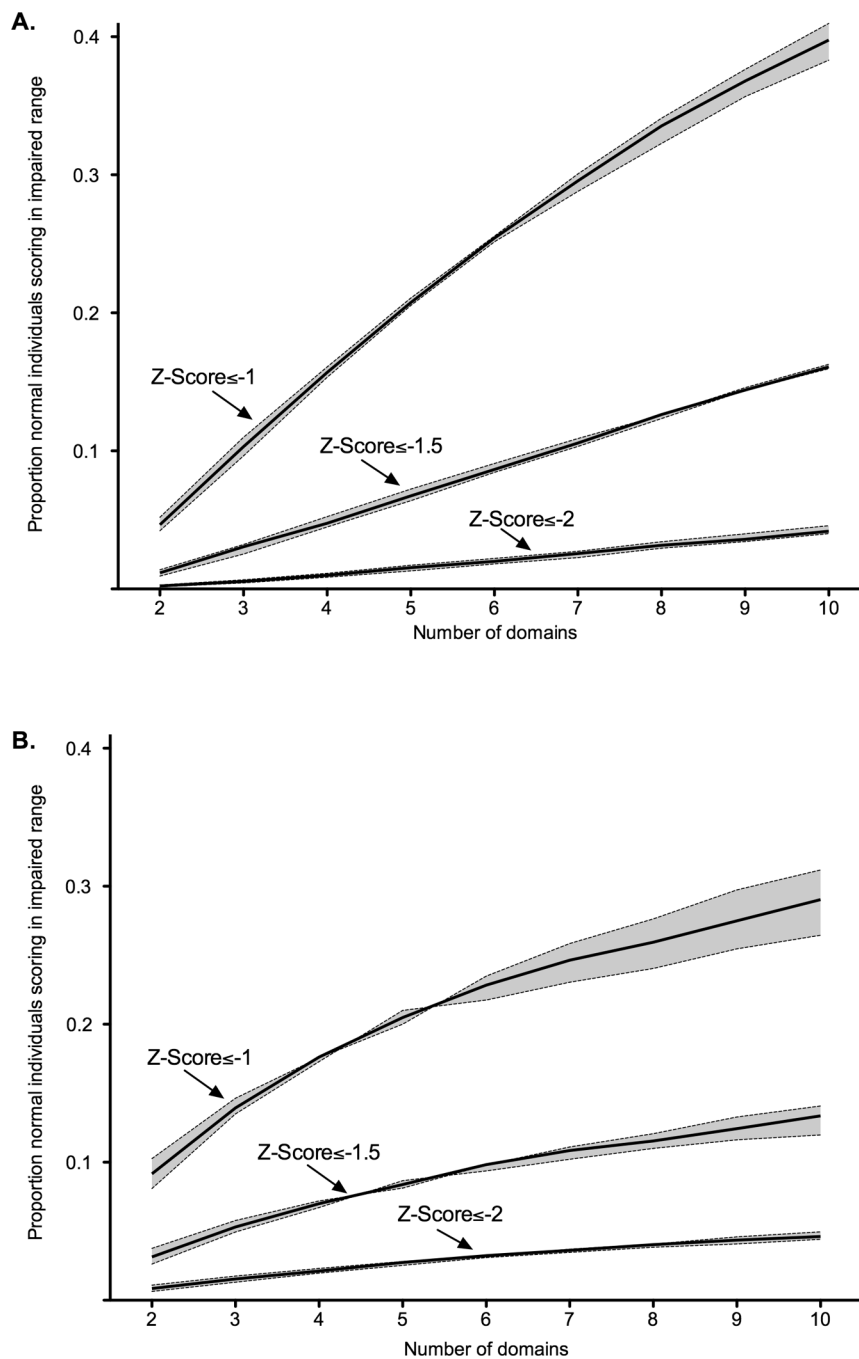


Figure 2.

The proportion of a simulated normal population who meet criteria for cognitive impairment using a variable number of domains and tests per domain.* (A) Correlation structure derived from actual Kenyan sample (B) Moderate correlation.

* The solid line represents the false-positive rate obtained using three tests per domain. The shaded areas around the line represent the range of false-positive rates obtained when using between two and five tests per domain. Simulations assumed correlated normal distributions around a true average score ($SD=0$). The average test score across a domain defined an

abnormal domain. Z-scores were truncated at ± 3 . Simulations were performed using 100,000 repetitions leading to a Monte Carlo error of approximately ± 0.2 – 0.4% .

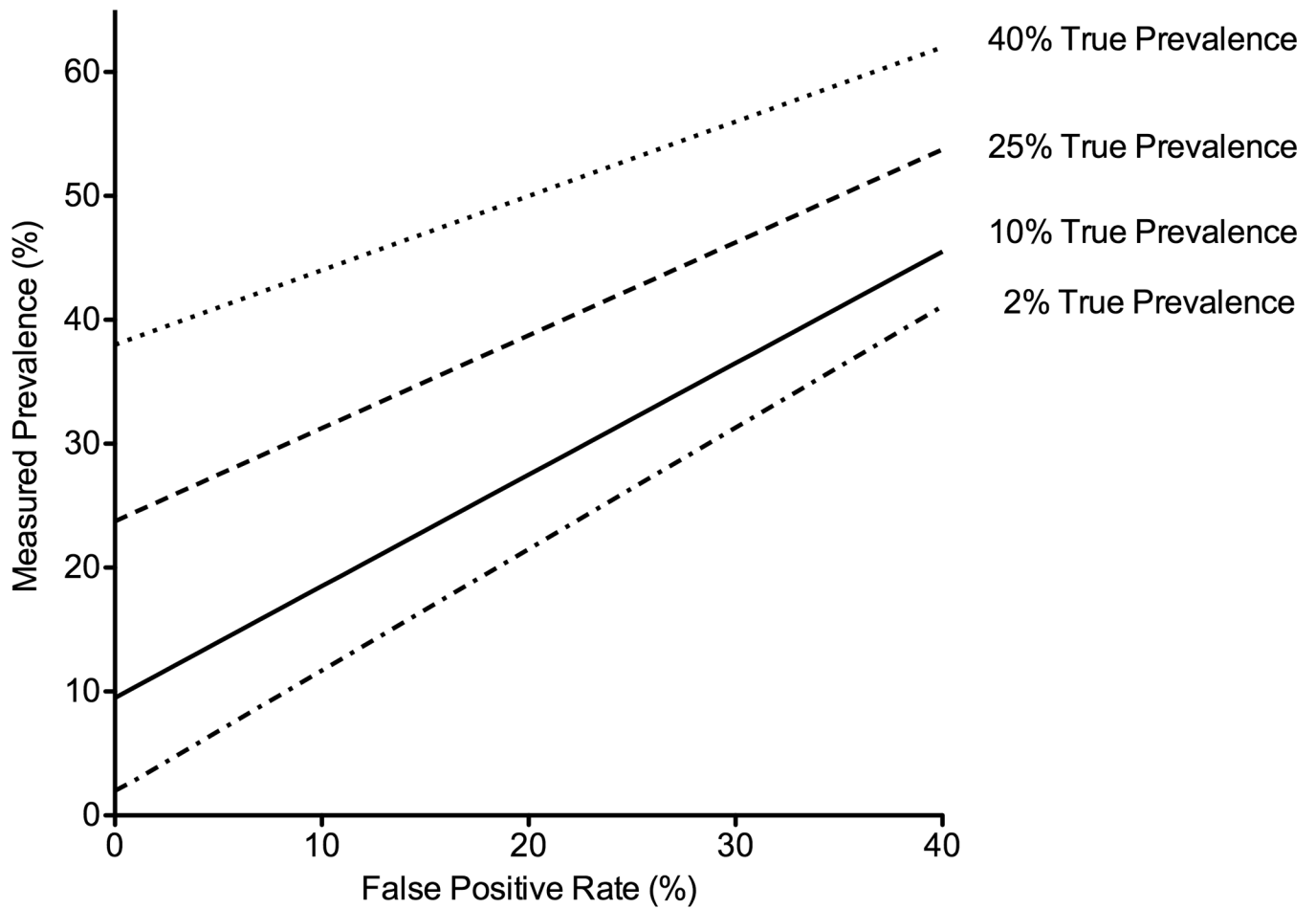


Figure 3. Measured prevalence as a function of the false-positive rate and true prevalence.*
*Assumes a false negative rate of 5%.

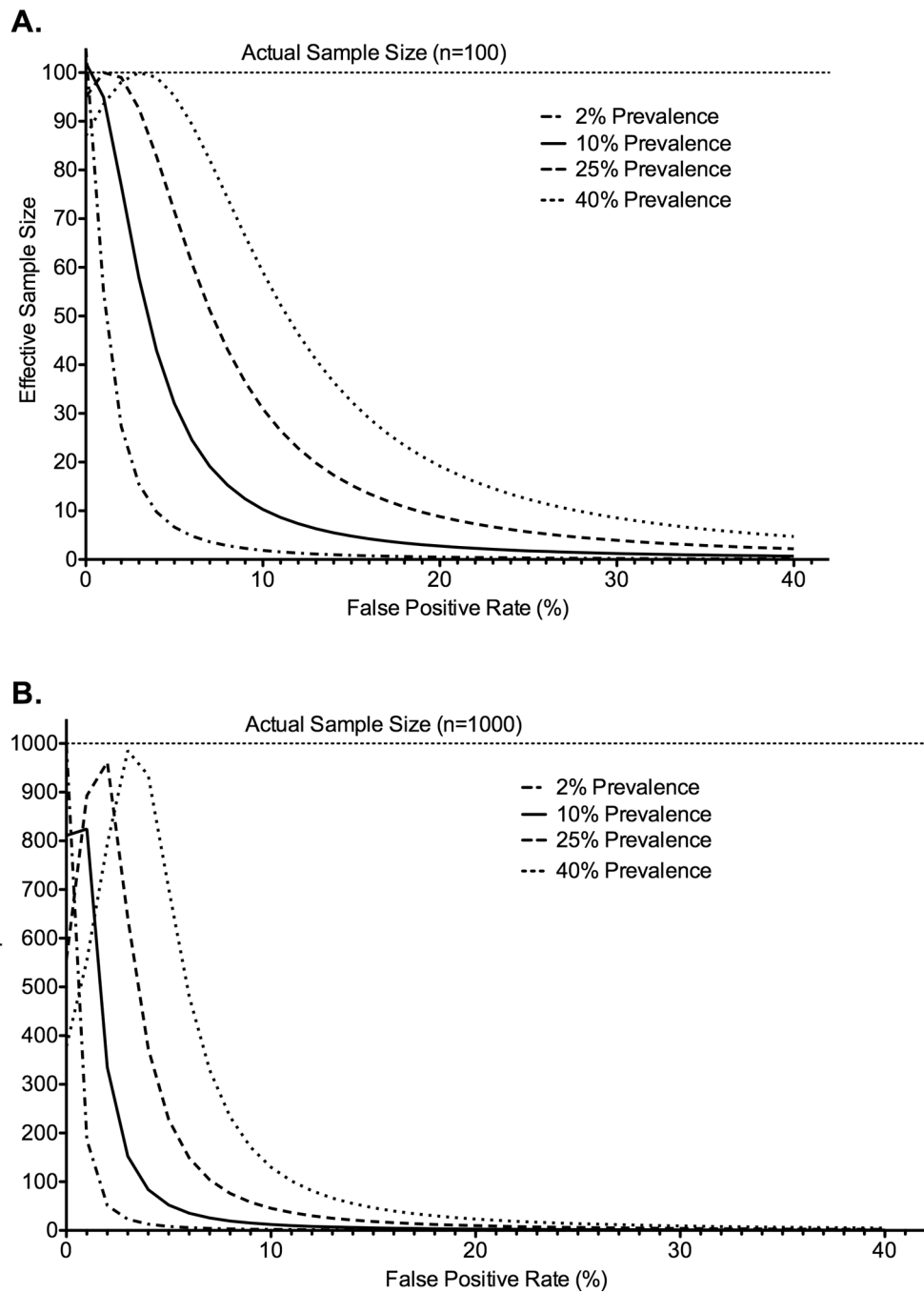


Figure 4. Effective sample size as a function of the false-positive rate, true prevalence of disease and actual sample size. A. Sample size n=100. B. Sample Size n=1000.*

*Assumes a false negative rate of 5%.