Number of Negative Tests Required to Exclude Specified Levels of Agent Abuse in Equine Forensic Chemistry

by

Nathan C. Hall, Jyan-Ming Yang, Scott D. Stanley, George D. Mundy, Richard Kryscio, H.H. Tai, W. Allan Rees, Jerry W. Blake, W.E. Woods, and Thomas Tobin

Maxwell H. Gluck Equine Research Center
Departments of Veterinary Science,
Chemistry and Statistics
The School of Pharmacy
The Graduate Center for Toxicology
University of Kentucky
Lexington, Kentucky 40506
and
The Kentucky Racing Commission
Ironworks Pike, Lexington, Kentucky.

Published as Kentucky Agricultural Experiment Station Article #90-4-17 with the approval of the Dean and Director, College of Agriculture and Kentucky Agricultural Experiment Station.

Publication #174 from the Kentucky Equine Drug Testing and Research Programs, Department of Veterinary Science and the Graduate Center for Toxicology, University of Kentucky.

Supported by grants entitled "Development of a test for procaine in horses" and "Thresholds and clearance times for therapeutic medications in horses" funded by The Kentucky Racing Commission and The Equine Drug Council, Lexington, KY and by research support from the National Horseman's Benevolent and Protective Association, New Orleans, LA.

Correspondence should be addressed to Dr. Tobin, 108 Maxwell H. Gluck Equine Research Center, Department of Veterinary Science, University of Kentucky, Lexington, Kentucky 40546.

SUMMARY

The number of negative forensic tests that must be performed to ensure that a defined level of abuse of a specific agent can be excluded is currently undocumented. In this communication, we show that at least 500 consecutive negative tests must be run in order to be 99% confident that the rate of agent abuse is not greater than 1.0% in an infinite population. To increase the level of confidence to 99.9 or 99.99%, the number of consecutive negative tests required increases to 700 and 1,000 tests, respectively.

The average rate of detection in US racing of all Association of Racing Commissioners International Class 1 and 2 agents is about 0.1%. To be 99.0, 99.9, or 99.99% confident that an abuse problem does not exist for a particular agent at a rate of abuse not greater than 0.01% (1 in 10,000 samples tested), the number of consecutive negative tests required increases to 46K, 69K and 92K tests, respectively.

These data show that large numbers of tests must be run to confidently exclude the very low rates of specific substance abuse that racing seeks to enforce. This necessity to test a large proportion of the samples for effective control makes the cost effectiveness of testing methods critical. These results also suggest that testing strategies based on "intensive" inspection of reduced numbers of samples are unlikely to be cost effective or forensically productive.

INTRODUCTION

Central problems in equine forensic chemistry include deciding when to deploy specific tests, the conclusions that can be drawn from a series of negative tests, and, conversely, determining when it is appropriate to withdraw a specific test (Tobin, 1981). This problem has become more acute with the general availability of ELISA (Enzyme-Linked Immuno-sorbent Assay) tests (Tobin et al., 1988). ELISA tests are highly specific, generally detecting only the agent against which the test was raised and one or two structurally closely related agents. Deployment of an ELISA test, therefore, allows one to test for a single agent at relatively low cost but very high sensitivity (Woods et al., 1992). The question therefore arises as to how long one should continue to use a specific ELISA in the face of repeated negative results, and, more importantly, what conclusions can be drawn about the rate of agent abuse when a specific number of negative test results have accumulated.

In answering these questions, we have chosen to focus on control of Association of Racing Commissioners International Class 1, 2 and 3 agents; agents with the highest potential to influence the performance of horses and relatively low rates of abuse. For example, review of the detection rates for these agents developed several years ago (Woods et al., 1985) and more recently (Mundy et al., 1994) have shown that the rate of detection of this entire class of agents in post-race samples from racing horses runs at about one identification per 1,000 samples analyzed. These are very low rates of agent abuse, much lower, for example, than those observed in human athletics or human Drugs of Abuse (DOA) testing (Cowan, 1995). As will be seen from the statistical data presented, the need to detect and prevent very low (0.1 to 0.01%) rates of agent abuse in racing horses greatly increases the number of tests that must be performed.

Here c is the number of positive identifications in n biological fliud samples selected at random from an infinite population having a level of drug abuse P. Also,

RESULTS

The first question addressed concerns the practical conclusions that can be drawn when a certain number of consecutive negative test results have been obtained. The data of Figure 1 demonstrates the number of consecutive negative tests which must be obtained to establish, at varying levels of confidence, that the level of agent abuse is not greater than a specified level. The levels of agent abuse selected were, respectively, 1.0, 0.1 and 0.01% in an infinite population of horses. Figure 1 shows that in order to exclude a level of agent abuse of 1.0% or more at the 99% level of confidence, the number of consecutive negative tests which must be run is approximately 500. If exclusion of lower rates of abuse is required, the number of consecutive tests required increases in approximately ten fold increments, to 5,000 for exclusion levels of 0.1% and 50,000 for an exclusion level of 0.01%, respectively.

If we wish to exclude these rates of abuse at higher levels of confidence, the number of samples that must be found negative to attain this same level of exclusion increases. For example, to establish at a 99.99% level of confidence that agent abuse is not greater than 1.0, 0.1 or 0.01%, the number of consecutive negative tests required increases to 920, 9,200 and 92,000 tests, respectively.

These results were developed based on a model assuming an infinite population of samples. For finite populations, the actual population size influences the number of consecutive negative tests required to statistically prove (at the 99% confidence level) that the rate of agent abuse is less than 1.0, 0.1 and 0.01%, as presented in Figure 2.

Overall, as the population size increases, the fraction of the population that must be tested to exclude a given rate of agent abuse decreases. For example, for a population of 10,000 samples, fewer than 5% of samples must be tested to exclude abuse above 1%, about 35% to exclude abuse above the 0.1% rate, and essentially 100% of samples to exclude abuse at the very low rate of greater than 0.01% abuse (Figure 2).

On the other hand, if the population of samples is 100,000, all of these percentages drop significantly, with testing of 1% excluding abuse above 1%, testing of 5% excluding above 0.1% and testing of 40% of the samples excluding abuse above the 0.01% rate of abuse.

Figure 3 presents the rates of agent abuse that can be excluded at the 99% confidence limit when 0 negative tests are reported in sample populations ranging from 1,000 to 10,000. This figure assumes the sample populations are a subset of an infinite population. In the case of a sample population size of 1,000, if all of the samples are tested, one can be 99% confident that the rate of agent abuse is not greater than about 0.45% (a relatively high rate of abuse). To exclude an abuse rate of greater than 0.1% at the 99% confidence level, testing a sample population of at least 4,500 is necessary. To exclude an abuse rate of greater than 0.01% at this confidence level, testing a sample population of greater than 10,000 is necessary. Again, to confidently define the true rate of agent abuse at acceptable levels, testing of relatively large sample populations is necessary.

As the number of negative tests obtained increases, one can exclude lower and lower rates of agent abuse (Figure 3). As a general rule, the first 1,000 consecutive negative tests enables one to rule out levels of agent abuse above about 0.45%. However, approximately 4,500 consecutive negative tests enables one to rule out agent abuse greater than 0.1%, with higher numbers of sequential negative samples required to exclude lower rates of substance abuse.

Figure 4 shows the conclusions that can be drawn about the true level of agent abuse as sequential positives are obtained in a population of 10,000 samples. As the number of positives increases, the true level of agent abuse also increases. Because the 10,000 samples are only a sample of an infinite population, the true level of agent abuse is always greater than the observed level of agent abuse.

DISCUSSION

These results bear directly on the deployment of tests and interpretation of the result of equine testing. Review of the literature in this area shows that the rates of agent abuse varies from a high of between 12 and 20%, in what are effectively uncontrolled situations, to rates of abuse in controlled situations of less than 0.01% for specific agents. The goal of an effective equine forensic program is to ensure that the rate of medication abuse in racing remain at the very low (> 0.1%, overall) rates historically seen in racing. As pointed out in the introduction, these rates of agent abuse are exceptionally low, much closer to zero abuse rates than those found in related areas of human forensic testing (Cowan, 1995).

In the first possible circumstance, that of epidemic agent abuse, detection of abuse does not require a large number of tests. If the rate of agent abuse is 5% or greater, then detection essentially awaits deployment of an effective test, which yields dramatic results very quickly. For example, when ELISA tests were first deployed in the southwestern US, numerous agent identifications were made and patterns of agent abuse that had presumably essentially unbroken for the larger part of this century were terminated within a matter of months (Tobin et al. 1988).

On the other hand, if the rate of abuse is less than epidemic, then the results of introduction of a test may be less clear-cut. For example, a not uncommon approach to the introduction of ELISA testing was to run a series of 500 or so tests and, on the basis of 500 consecutive negative tests, conclude that abuse of the agent was not occurring. However, as shown in Figure 1, all that can be concluded from a series of 500 negative tests is that the rate of agent abuse is not greater than about 1.0% at the lowest level of confidence (99% confidence, UCL of 1.0%) that we have analyzed. If we want to be highly confident (99.99% confident) that abuse at a rate greater than 1.0% is not occurring, then we must test close to 1,000 samples.

In terms of the rates of agent abuse occurring in horse racing, any rate of abuse equal to or greater than 0.1% is a very high rate of abuse. For example, common patterns of abuse are likely to be occasional use of a new agent that is being "tried" by motivated individuals. Beyond this, our experience in horse racing has been that once an agent is identified and effective administrative action taken, the rate of agent abuse drops dramatically. For example, when anabolic steroid tests were first introduced in England, the rate of abuse of these agents dropped from about 12% to zero for two years. Similarly, there was a period of at least 12 months after the introduction of ELISA tests in the American Southwest during which these tests were not deployed. When these ELISA tests were again deployed after a one year interruption, we were surprised to find that agents that had previously been widely used but had not been tested for one year (oxymorphone, buprenorphine and sufentanil) were not being used. The epidemiology of agent abuse in racing appears to be that once an agent is "called", horsemen are very reluctant to tempt fate (or the authorities) and use this substance again. Re-use does occur, but, initially at least, apparently at very low rates of use. This low rate of re-introduction is likely broadly similar to the rate of introduction of new agents, for example, the recent "introductions" of viloxazine and romifedene in California racing (Stanley, 1995).

The goal of an effective testing system is therefore to keep the level of abuse of both well-characterized and novel agents at very low rates of use, less than about one identification per 1,000 samples, the average rate at which identifications of agents of abuse have been made in North American racing over the last fifteen years. Since these identifications generally involve a number of agents, it is reasonable to conclude that a good "not to be exceeded abuse rate" for individual agents in racing horses is to exclude rates of agent abuse of not greater than 0.01% for any given agent. This is a considerable technical challenge, and to exclude these low levels of agent abuse, substantial fractions of most populations must be tested.

Figure 2 shows the effect of population size on the total number of samples that must be tested to exclude a given rate of agent abuse at the 99% confidence level. If the size of the population tested is about 1,000 samples, we must test about 40% of the samples to exclude rates of agent abuse of greater than 1.0%. However, if we want to increase the level of agent abuse that we can exclude to 0.1%, or one sample in 1,000, we must increase the fraction of samples tested to nearly 100% of all samples. Finally, if we want to raise the level of agent abuse that can be excluded to not greater that 0.01%, then we must test a minimum population size of 10,000 samples (testing approximately 100% of these) or about 40% of a population of 100,000 samples. These are significant numbers of samples to test. These figures make clear the basic message of this paper: to exclude very low rates of agent abuse, one must test very substantial fractions of the total population of samples presented for testing.

This need to test very substantial fractions of the population of samples presented for testing, combined with the wide scope of equine forensic testing, makes the cost effectiveness of the testing methods critical. To effectively control abuse of an agent at the low levels of abuse traditional in the racing industry, a test must be applied to a large proportion of the samples presented to the analytical laboratory in each year. If the test is expensive, and can be applied only to a small fraction (say 10%) of the samples presented to the laboratory, its deployment will only be able to effectively exclude rates of agent abuse between 1.0% and 0.1%, not an effective performance level by the standards required in equine forensic testing. Deployment of such tests is easiest to justify when there is evidence to suggest high rates of abuse of these specific agents.

These results also raise significant questions about the philosophical approach of intensive screening of reduced numbers of samples, as suggested by the Jockey Club/McKinsey report (1991). Reducing the number of samples presented to test has the obvious effect of reducing the probability of presenting a sample containing an abused agent to the test. If a drug is not widely abused, as most class 1 or class 2 agents are not, then reducing the number of samples tested equivalently reduces the level of agent abuse which can be excluded. By the same argument, if the test used at a reduced rate is more expensive than other testing approaches, the cost efficacy of this test must be measured as a function of the increased cost multiplied by the reciprocal of the fractional use of the reduced testing efficacy.

In summary, therefore, to insure that the low rates of agent abuse demanded by the racing industry are assured, a substantial fraction of the samples submitted for testing must be analyzed. This report also shows that conclusions can be drawn about the rates of drug abuse in the population based on the number of consecutive negative tests reported, and sets forth the mathematical basis for these conclusions. It is clear that any reduction in the number of samples analyzed reduces the efficacy of testing and the level of drug use that can be excluded. The best testing methods should be inexpensive and applicable to a large fraction of the samples tested;

techniques that can only be applied to a fraction of the samples presented for testing are correspondingly less effective forensic techniques.

References

Building a world-class drug detection system for the racing industry: A national strategic plan. Final Report, McKinsey & Company, Inc. (1991).

Beyer, W. H. (ed) Handbook of Tables for Probability and Statistics 2nd Edition, Chemical Rubber Company, Cleveland, Ohio, p. 219 (1968).

Cowan, D. An overview of how drug testing programs outside of racing handle the question of threshold/limitations. in Testing for Therapeutic Medications, Environmental and Dietary Substances in Racing Horses, Proceedings of Workshop. Tobin, T., Mundy, G. D., Stanley, S. D., Sams, R. A., and Crone, D. (eds), Lexington, KY, p. 49. August, 1994. Published 1995.

Mundy, G. D. and Stanley, S. D. Elective testing. in Testing for Therapeutic Medications, Environmental and Dietary Substances in Racing Horses, Proceedings of Workshop. Tobin, T., Mundy, G. D., Stanley, S. D., Sams, R. A., and Crone, D. (eds), Lexington, KY, p. 119. August, 1994. Published 1995.

Stanley, S. D. Personal Communication. (1995).

Tobin, T. Drugs and the Performance Horse. Thomas, C. C., Springfield, IL. (1981).

Tobin, T., Watt, D. S., Kwiatkowski, S., Tai H.-H., and Blake, J. W. Non-isotopic immunoassay drug tests in racing horses: A review of their application to pre- and post-race testing, drug quantitation, and human drug testing. *Res. Comm. Chem. Pathol. Pharmacol.*, 62(3): 371-395, (1988).

Woods, W. E., Chay, S., Houston, T., Blake, J. W., and Tobin, T. Efficacy of testing for illegal medication in horses. *J. Am. Vet. Med. Assoc.*, 187: 927-930, (1985).

Woods, W.E., Stanley, D. S., Yang, J-M, Cook, D., Gerkin, D. F., Detra, R. L., Sams, R. A., Jeganathan, A., Watt, D. S., Tai, H-H., and Tobin, T. ELISA detection of azaperone in the racehorse. in Short, C. R., (ed), Proc. 9th Intl'n Conf. Racing Analysts Vets., New Orleans, LA, 187-192, (1992).

Figure 1. Effect of specified level of agent abuse and desired confidence level on the number of negative tests required.

The solid lines show the relationships between the number of consecutive negative tests required and the desired level of confidence establish rates of agent abuse less than 1.0%, 0.1% and 0.01% respectively. This analysis assumes an infinite population size.

Missing

Figure 2. Relationship between population size, percent of population sampled and rates of agent abuse excluded.

The solid lines show the relationships between population size and percentage of the population that must be sampled to establish that the rates of agent abuse are less than 1.0%, 0.1% and 0.01%, with the level of confidence set at 95%.

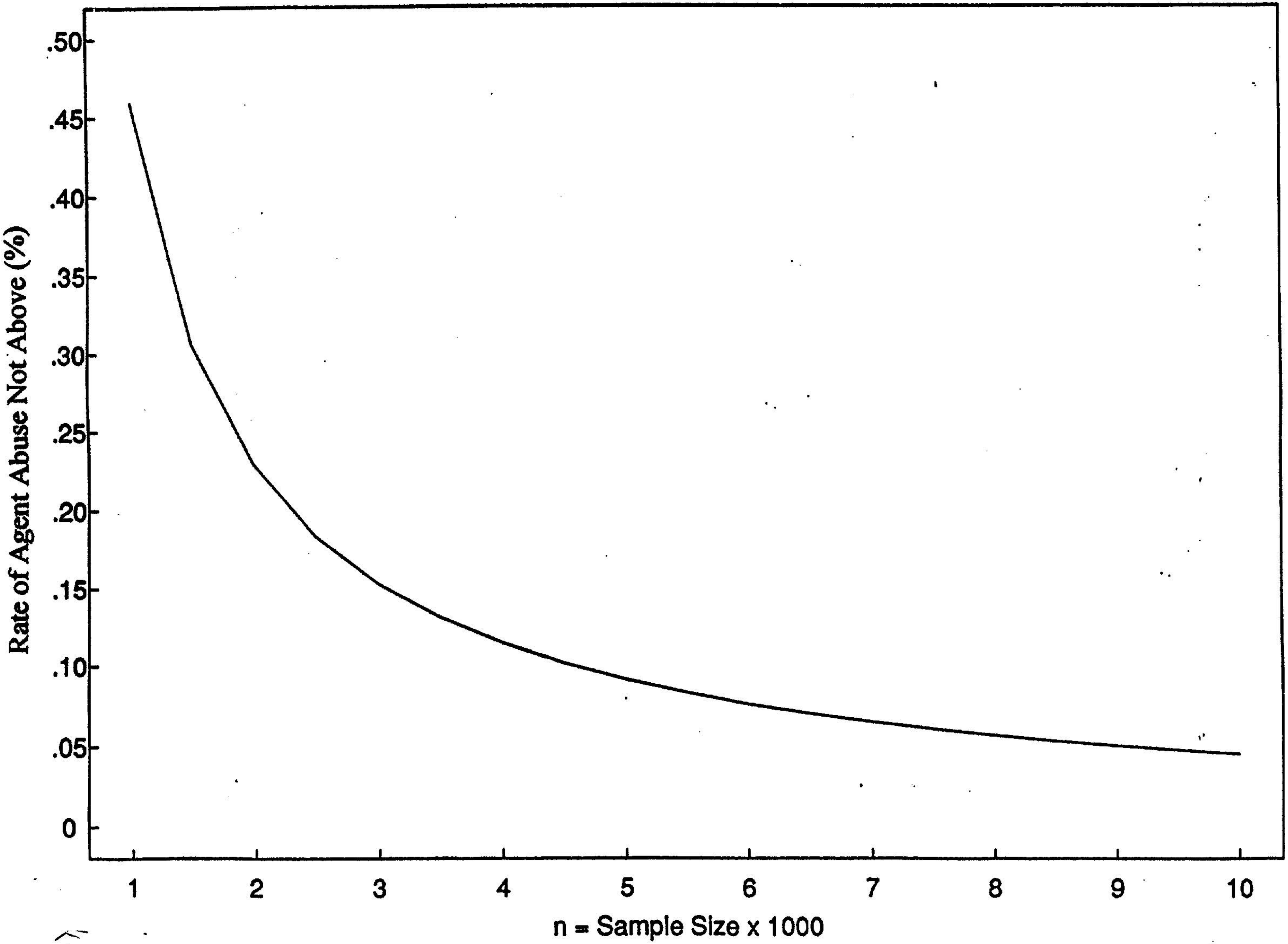
Figure 3. Level of agent abuse when zero positives are found among n samples.

The solid line shows the relationship between the rate of agent abuse (vertical axis) against the number of sequential negative tests observed (horizontal axis). The population size is assumed to be infinite and the statistical level of confidence is 99%.

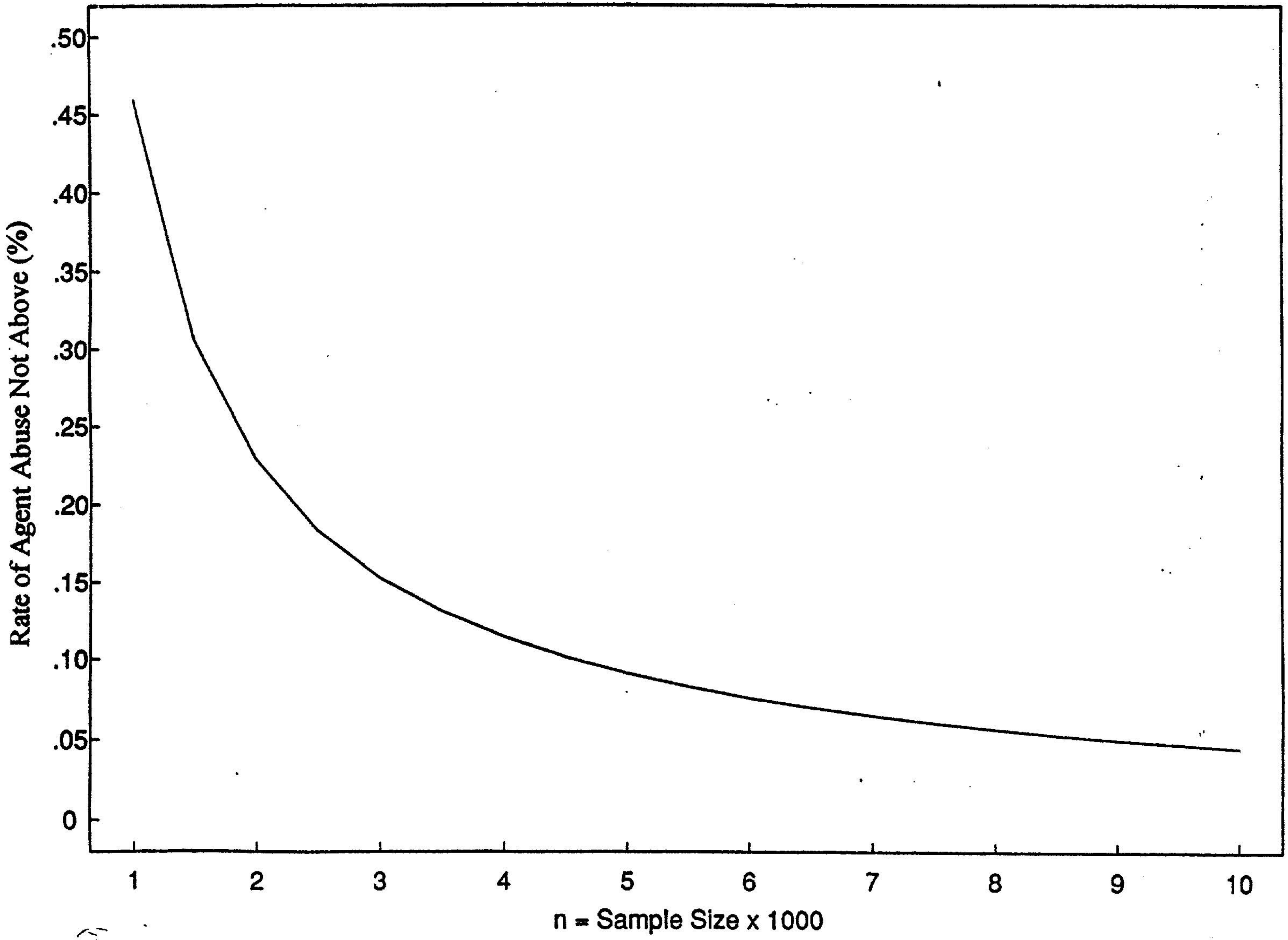
Figure 4. True level of agent abuse when x positive tests are obtained from a sample of 10,000.

The solid circles (•—•) show the relationship between the true rate of agent abuse (vertical axis) and the number of positive tests reported in 10,000 samples. The open circles (o—o) show the rate of agent abuse observed strictly from the samples tested. This analysis assumes a sample size of 10,000 from an infinite population, with the level of confidence being 99%.

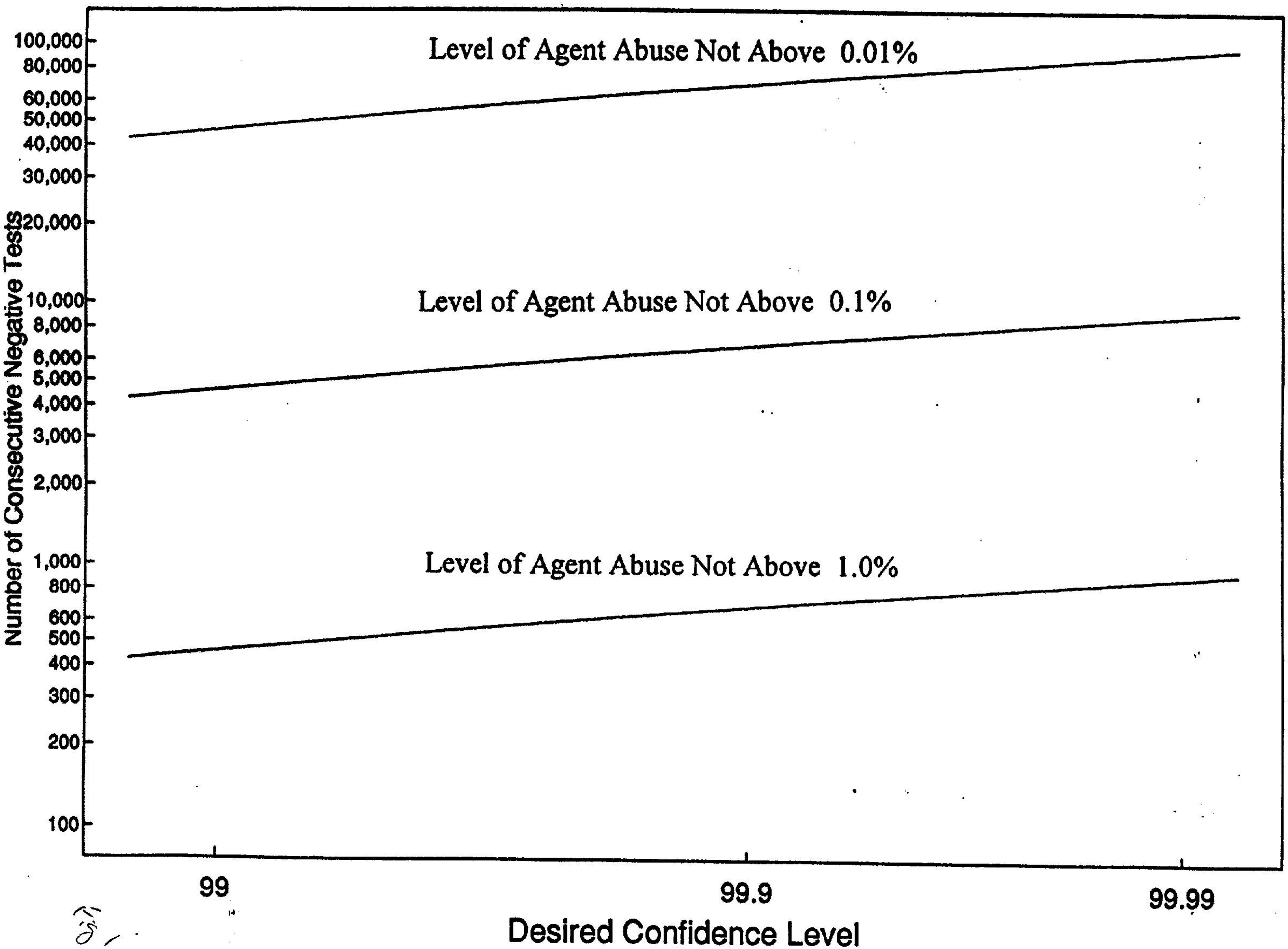
99% Upper Confidence Limit on True Level of Agent Abuse
When 0 Positives are Found Among n Tests



99% Upper Confidence Limit on True Level of Agent Abuse
When 0 Positives are Found Among n Tests



Number of Samples Required to Exclude a Specific Level of Agent Abuse



99% Upper Confidence Limit for Levels of Agent Abuse
When x Positive Tests are Obtained From 10,000 Samples