Checked by WRW
03/17/04

Friday, February 27, 2004
3:00 pm

Title: Number of Negative Tests Required to Exclude Specified Levels of Agent Abuse in Equine Forensic Science

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

Address: 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 Horse Racing Authority 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 #349 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 drug/medication/agent can be confidently excluded is currently undocumented. In this communication, it is shown that about 450 consecutive negative tests must be obtained to be 99% confident that the rate of agent abuse is not greater than 1.0% in an infinite sample population. To increase the level of confidence to 99.9 or 99.99%, the number of consecutive negative tests required increases to about 690 and 920 tests, respectively.

In United States horse racing the average rate of detection of all Association of Racing Commissioners International Class l, 2 and 3 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 is 45,000, 69,000 and 92,000 tests, respectively.

These calculations show that relatively large numbers of tests must be run to confidently exclude the low rates of specific substance abuse that racing authorities seek to maintain. This necessity to test a large proportion of the sample population for effective control makes the cost effectiveness of testing methodologies critical. These results may also suggest that testing strategies based on intensive inspection of reduced numbers of samples are less likely to be both cost effective and/or forensically productive.

INTRODUCTION

Three central problems in equine forensic chemistry are deciding when to deploy specific tests, the conclusions that can be drawn from a sequence of consecutive negative tests, and, conversely, determining when it is appropriate to withdraw or terminate application of a specific test.[1] This problem has become more acute with the general availability of ELISA (Enzyme-Linked Immuno-Sorbent Assay) tests [2] because they are highly specific, generally detecting only the agent against which the test was raised and one or, rarely, more structurally related agents. Therefore, an ELISA test allows for the testing of a single agent, or a relatively small number of agents, at relatively low cost but at very high sensitivity and with a high level of confidence [3]. The questions therefore arise as to how long a specific ELISA test should continue to be used in the face of repeated negative results and what conclusions can be drawn about the rate of agent "abuse" when certain specific numbers of consecutive negative test results have accumulated.

In answering these questions, it is instructive to focus on control of Association of Racing Commissioners International Class 1, 2 and 3 agents, those with the highest potential to influence the performance of horses and relatively low rates of abuse. For example, reviews have shown that the rate of detection of these three classes of agents (~550 agents in total) in post-race samples from racing horses is about one identification per 1,000 samples analyzed [4,5]. These are very low rates of agent abuse, much lower than those observed in human athletics, human Olympics, or human Drugs of Abuse (DOA) testing.[6] Additionally, 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 requires high numbers of tests.

A second question that arises is how to optimize the cost effectiveness of equine drug testing. If a laboratory has to choose between inexpensive tests that are applicable to virtually all samples and more expensive, elaborate tests that can only be applied to a much smaller proportion of the samples, the question ~~arises as to~~ is which strategy should be employed. This approach was set forth in the McKinsey report,[7] which recommended reducing the number of samples tested but applying more "in-depth" testing to those samples. In this report, this approach to equine forensic testing is evaluated.

METHODS

Statistical calculations were performed by one of us (Kryscio) to determine the minimum number of samples needed to be tested to conclude, at different levels of confidence, that the rate of agent abuse does not exceed a specified percentage. The analysis assumes that a negative analytical result is 100% accurate for the agent in question at the effective limit of detection (LOD) of the test. While any test can give false positive results at a given LOD, ELISA tests, when correctly performed, rarely, if ever, yield false negative results.

The following equation, taken from the Handbook of Tables for Probability and Statistics[8], was used in most calculations:

$$\sum_{x=0}^{c}\binom{n}{x}P^{x}(1-P)=\alpha$$

Here c is the number of positive identifications in n biological fluid samples selected at random from an infinite population having a level of drug abuse P. ~~Also, α reflects the level of confidence, which is set at $100 * (1-\alpha)$ percent.~~

The level of confidence is set at $100*(1-\alpha)$ percent.

RESULTS

Figure 1 shows the number of consecutive negative tests required to establish, at varying levels of confidence, that the level of agent abuse is not greater than a specified level (1.0%, 0.1% and 0.01%). To exclude a level of agent abuse of 1.0% or more, at the 99% confidence level, the required number of consecutive negative tests is approximately 450. If exclusion of lower rates of abuse are required, the number of required consecutive tests increases in approximately ten fold increments, to 4,500 for exclusion levels of 0.1% and 45,000 for an exclusion level of 0.01%. Likewise, higher levels of confidence require even more consecutive negative test samples. For example, a 99.99% level of confidence that agent abuse is not greater than 1.0%, 0.1% or 0.01% requires 920, 9,200 and 92,000 consecutive negative tests, respectively.

Figure 1 was developed using a model that assumed an infinite population of samples. For finite populations, the actual population size influences the number of consecutive negative tests required to statistically prove that the rate of agent abuse is not above 1.0, 0.1 and 0.01% (Figure 2). As the population size decreases, the proportion of the population that must be tested to exclude a given rate of abuse increases.

the analyst by using that substance again. Reappearance of previously detected drugs does occur but, initially at least, at very low rates. This low rate of reappearance is likely similar to the rate of introduction of new agents, for example viloxazine and romifedene, described in California racing in the mid-nineties[1].

The goal of an effective equine drug and medication testing system is to ensure that the level of use of both well-characterized and novel prohibited agents does not exceed very low rates of use, hopefully less than about one identification per 1,000 samples (0.1%), the average overall rate at which identifications of abuse agents have been made in North American racing over the last fifteen years. Since these identifications generally involve a number of different agents, it is reasonable to conclude that a good not-to-be-exceeded abuse rate for individual agents is not greater than 0.01%. This is a considerable technical challenge, and to insure these very low levels of agent abuse, substantial fractions of most sample 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 abuse at the 99% confidence level. If the size of the population tested is 1,000 samples, then ~35% of the samples must be tested to exclude rates of abuse greater than 1.0%. However, to increase the level of abuse ~~that can be excluded to~~ not above 0.1%, or one "positive" in 1,000 samples, almost 100% of the samples would have to be tested. Finally, to raise the level of abuse that can be excluded to not greater than 0.01% would require a population size of 10,000, and 100% of the samples would have to be tested. If the population size were increased to 100,000, about 40% of the population would have to be tested. These are substantial numbers to test. The figures make clear the basic message of this paper: to insure very low rates of agent abuse, very large fractions of the total sample population must be tested.

This need to test large fractions of the sample population, combined with the wide scope of equine forensic testing, makes the cost effectiveness of the testing methods, and particularly the screening methods, critical. To effectively control abuse of an agent at the low levels traditional in the racing industry, effective tests 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 to only a small fraction (say 3.3%) of the samples presented, its use will only be able to effectively exclude rates of agent abuse between 1.0% and 0.1%, an ineffective performance level by equine forensic testing standards. Justification of such tests is easier when there is evidence to suggest high rates of abuse of these specific agents.

These results may raise significant questions about the philosophical approach of intensive screening of reduced numbers of samples, labeled "super testing" by the Jockey Club/McKinsey report.[7] Reducing the number of samples has the obvious effect of reducing the probability of testing a sample containing an abused agent. If a drug is not widely abused, as is the case with most class 1 or class 2 agents, then reducing the number of samples tested essentially reduces the level of abuse that 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, to insure that the low rates of agent abuse demanded by the racing industry are maintained, a substantial fraction of the samples collected for testing must be analyzed. Furthermore, this report 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

[1] Personal communication, S. Stanley, 1995

For example, for a population of 10,000 samples, ~5% of the samples must be tested to exclude abuse above 1%, ~40% must be tested to exclude abuse above the 0.1% rate, and 100% of samples must be tested to exclude abuse above the 0.01% rate (Figure 2). However, if the sample population is larger, for example 100,000 samples, the percentages required for testing drop significantly to 1%, 5%, and 40% of the population to exclude an abuse rate above 1%, 0.1% and 0.01%, respectively.

Figure 3 shows the conclusions that can be drawn about the true level of agent abuse as 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 fraction of an infinite population, the true level of agent abuse is always greater than the observed level of agent abuse.

DISCUSSION

These findings bear directly on the deployment of tests and the interpretation of equine drug testing results. Review of the literature shows that the abuse rates vary from a high of 12-20%, in what are effectively uncontrolled situations, to rates of abuse in controlled situations of less than 0.01%. The goal of an effective equine forensic program is to ensure that the rate of medication abuse remain at the very low rates (>0.1% to >0.01%) 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 drugs of abuse or workplace forensic testing.[6] if it's a superscript put it outside period (6). OR .(6)

In the worst possible circumstance, that of epidemic agent abuse, detection of that abuse does not require a large number of tests. If the abuse rate is 5% or greater, then detection essentially awaits deployment of an effective test, which will yield dramatic results very quickly. For example, when ELISA tests were first deployed in the southwestern US in the late 1980s, numerous agent identifications were made and patterns of abuse that presumably had been present for most of the 20th century were terminated within a matter of months [2].

On the other hand, if the rate of abuse is less than epidemic, the effect of introduction of a new test may be less clear-cut. For example, a not uncommon approach for ELISA testing was to run a series of 200 to 300 tests and, on the basis of 300 consecutive negative tests, conclude that abuse of the agent was not occurring. However, as shown in Figure 1, 450 consecutive negative tests are required to conclude at a 99% confidence level that the rate of abuse is not greater than 1.0%, an unacceptably high rate of abuse in equine forensic testing. To be 99.99% confident that abuse at a rate greater than 1.0% is not occurring, then ~920 samples must be tested. Any rate greater than 0.1% is a very high rate of abuse.

A common pattern of abuse consists of a new agent being "tried" occasionally by motivated individuals. Once that agent is identified (reported "positive") and effective administrative action is 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 abruptly dropped from about 12% to zero in two years. Similarly, after ELISA testing had been introduced into equine forensic science, there was a period of at least 12 months when these tests were discontinued in the American Southwest. When the ELISA tests were reintroduced after at least a one-year interruption, it was surprising to find that previously used agents (oxymorphone, buprenorphine and sufentanil) were not being abused. The epidemiology of agent abuse in racing appears to be that once an agent is "called", horsemen are reluctant to tempt either fate or

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 reduced fraction of the samples presented are correspondingly less effective in the overall context of achieving effective control of the misuse of performance-enhancing substances in racing horses.

Figure 1. Number of consecutive negative tests required to excluded specific levels of agent abuse at different confidence levels, assuming an infinite population.

The lower solid line shows the relationships between the number of consecutive negative tests required to excluded levels of agent abuse not above 1% at the 99% [square] 99.9% [triangle] and the 99.99% [circle] levels of confidence, respectively. The middle dashed line shows the relationships between the number of consecutive negative tests required to excluded at levels of agent abuse not above 0.1% at the 99% [square] 99.9% [triangle] and the 99.99% [circle] levels of confidence, respectively. The upper dotted line shows the relationships between the number of consecutive negative tests required to excluded at level of agent abuse not above 0.01% at the 99% [square] 99.9% [triangle] and the 99.99% [circle] levels of confidence, respectively.

Figure 2. Relationships between population size and percentage of the population that must be sampled to establish that the rates of agent abuse are less than certain specified percentages (1.0%, 0.1%, and 0.01%).

Figure 3. Relationship between the "true" level of agent abuse when "n" positive tests are observed from 10,000 samples drawn from an infinite population .

In this figure the rate of agent abuse excluded is presented on the vertical axis against the number of positive tests reported on the horizontal axis. The solid circles (•—•) represent 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%.

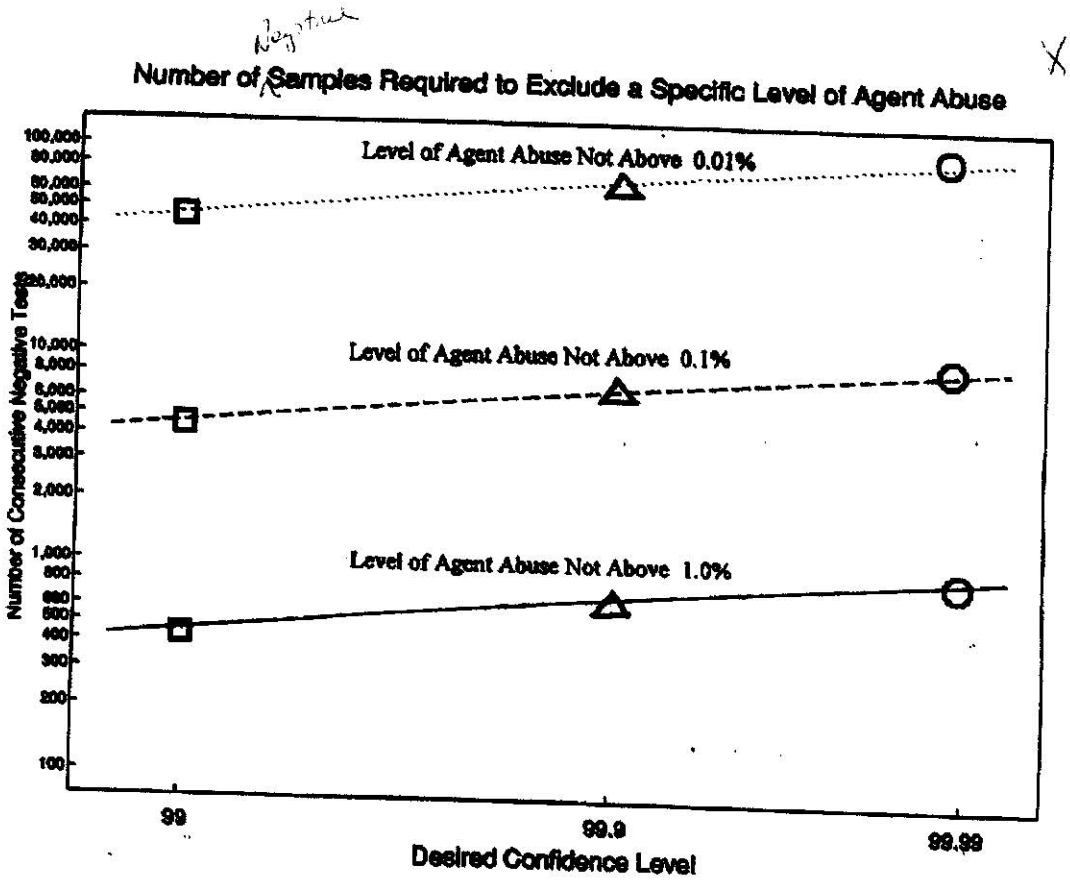

Number of Samples Required to Exclude a Specific Level of Agent Abuse
Level of Agent Abuse Not Above 0.01%
Level of Agent Abuse Not Above 0.1%
Level of Agent Abuse Not Above 1.0%
Number of Consecutive Negative Tests
100,000
80,000
60,000
50,000
40,000
30,000
20,000
10,000
8,000
6,000
5,000
4,000
3,000
2,000
1,000
800
600
500
400
300
200
100
99
99.9
99.99
Desired Confidence Level

Figure 1

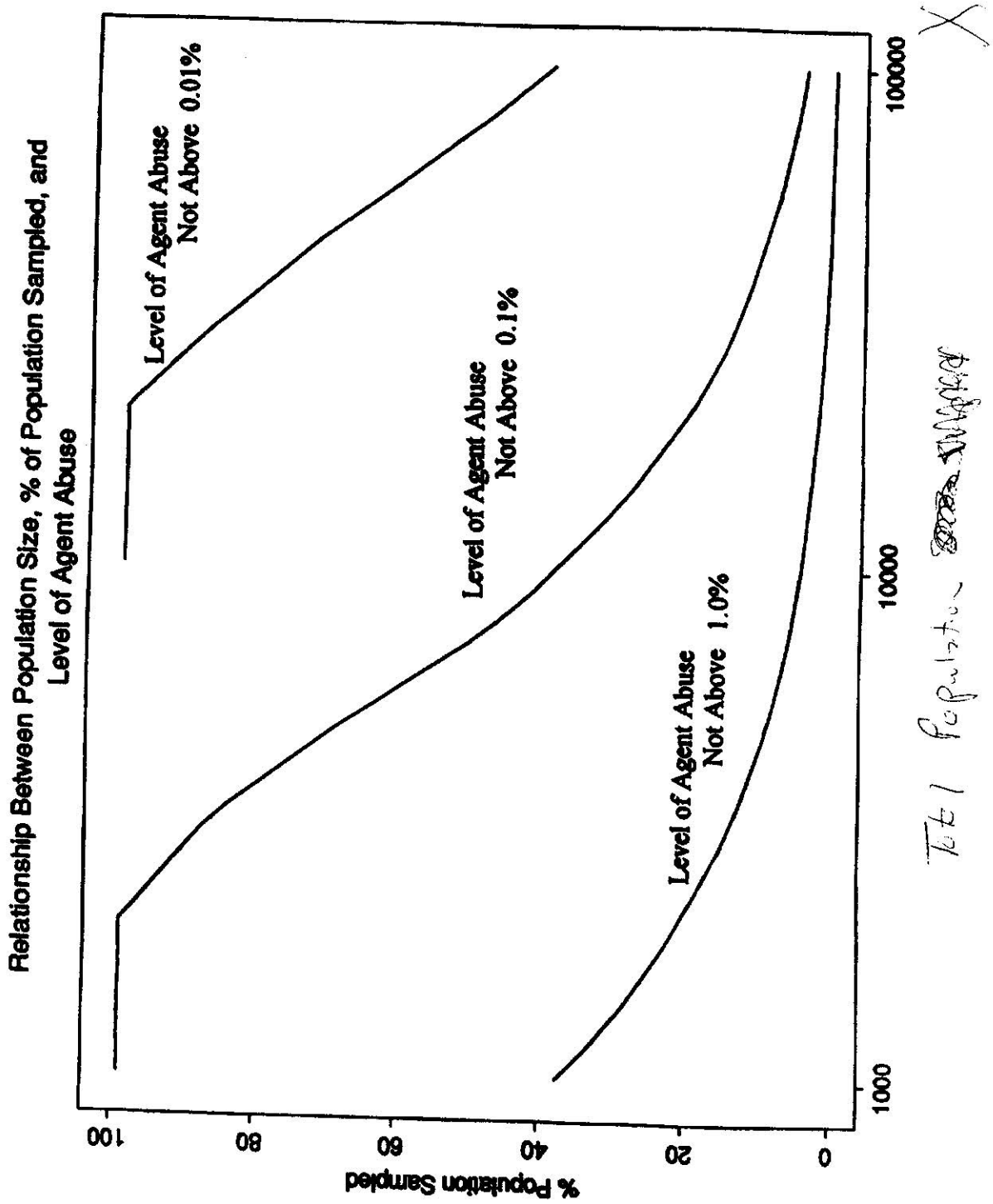
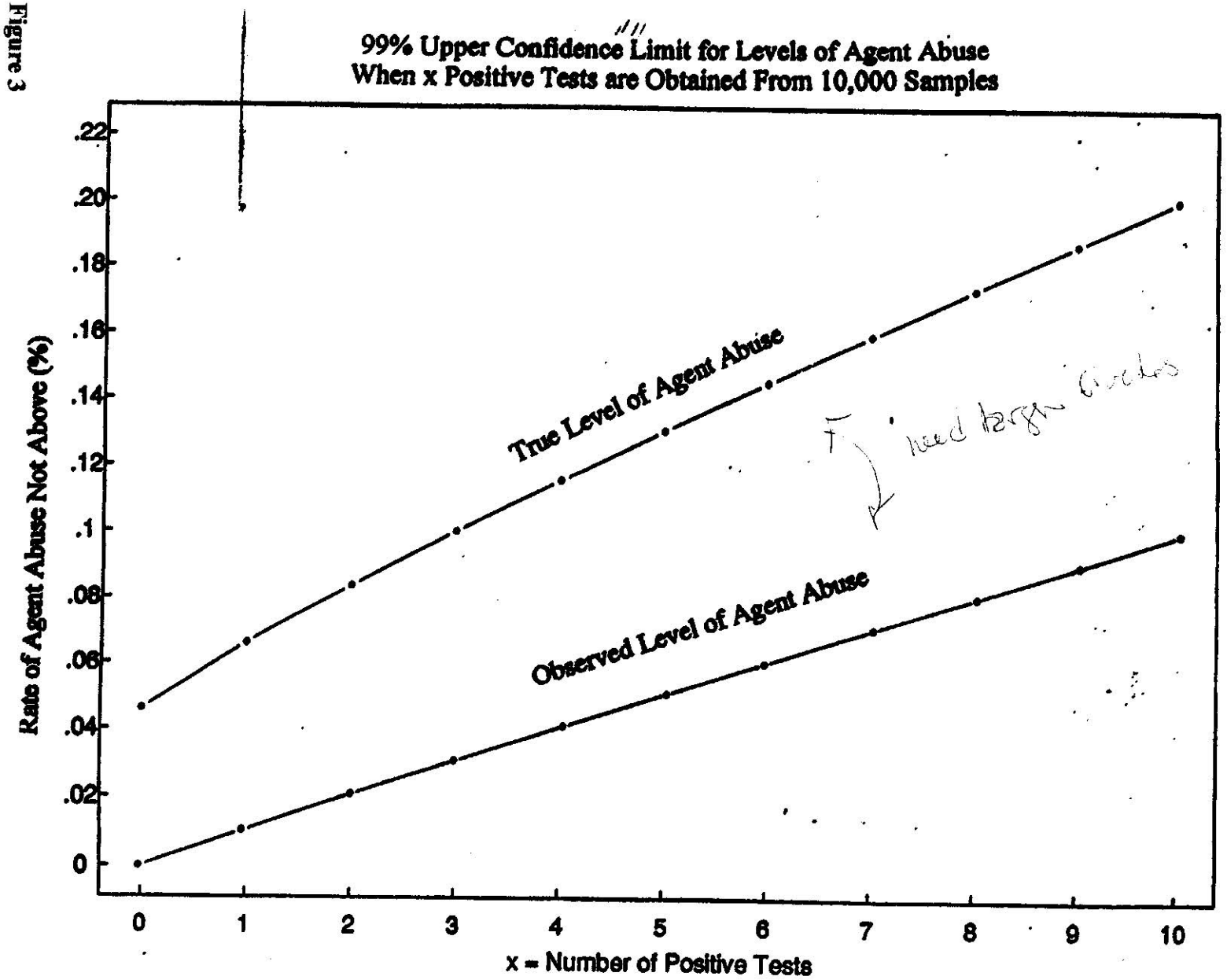

Relationship Between Population Size, % of Population Sampled, and Level of Agent Abuse
Level of Agent Abuse Not Above 0.01%
Level of Agent Abuse Not Above 0.1%
Level of Agent Abuse Not Above 1.0%
% Population Sampled
0
20
40
60
80
100
1000
10000
100000

Figure 2

Figure 3


99% Upper Confidence Limit for Levels of Agent Abuse
When x Positive Tests are Obtained From 10,000 Samples
Rate of Agent Abuse Not Above (%)
.22
.20
.18
.16
.14
.12
.1
.08
.06
.04
.02
0
True Level of Agent Abuse
Observed Level of Agent Abuse
need larger circles
0
1
2
3
4
5
6
7
8
9
10
x = Number of Positive Tests

Reference List

1. Tobin,T. Drugs and the Performance Horse 1981.

2. Tobin,T, Watt,D, Kwiatkowski,S, Tai,H, and Blake,JW. 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. Research Communications in Chemical Pathology and Pharmacology 1988; 62:371-395.

3. Woods,WE, Stanley,SD, Yang,JM, Cook,D, Gerkin,D, Detra,R, Sams,RA, Jeganathan,A, Watt,DS, Tai,HH, and Tobin,T. ELISA detection of azaperone in the racehorse. Proceedings of the International Conference of Racing Analysts and Veterinarians 1992; 9:187-192.

4. Woods,WE, Chay,S, Houston,T, and Blake,JW. Efficacy of testing for illegal medication in horses. Journal of the American Veterinary Medical Association 1985; 187:927-930.

5. Mundy,GD and Stanley,SD. Elective testing 1995.

6. Cowan,D. An overview of how drug testing programs outside of racing handle the question of threshold/limitations 1995:49.

7. The McKinsey Report. Building a World-Class Drug Detection System for the Racing Industry: A National Strategic Plan 1991:1-26.

8. Beyer,WH. Handbook of Tables for Probability and Statistics 1968; 2:219.