IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND
NORTHERN DIVISION

ORIGINAL

In re:	
UNITED STATES OF AMERICA,	Criminal No. 00-946PWG
v.	
	November 19, 2001
ERIC HORN,	
	Baltimore, Maryland
Defendant.	

EVIDENTIARY HEARING
BEFORE THE HONORABLE PAUL W. GRIMM
UNITED STATES MAGISTRATE JUDGE

APPEARANCES:

For the Government:

PAUL MONROE, Esq.
Special Assistant U.S. Attorney
U.S. Army Garrison
Building 310, Wing 10
Aberdeen Proving Grounds
Aberdeen, Maryland 21001

For the Defendant:

SASHA NATAPOFF, Esq.
Federal Public Defender's Office
Tower II, Suite 1100
100 South Charles St.
Baltimore, MD 21201

Audio Operator: Unknown

Transcription Company:

CompuScribe
9537 Elvis Lane
Seabrook, Maryland 20706
301/577-5882

Proceedings recorded by electronic sound recording, transcript produced by transcription service.

I N D E X

KEYNOTE: "---" denotes inaudible in the transcript.

and just bear with us, doctor. I am trying to make the most I can out of the time we have available.

MS. NATAPOFF: Your Honor, I take it that you have in front of you --

THE COURT: I do.

MS. NATAPOFF: -- all the Government's past submissions. Dr. Cole is going to be discussing some tables in particular, and I took the liberty of copying them in a packet. They are simply reproductions of what is already in.

THE COURT: Great.

MS. NATAPOFF: But I thought it may save some time.

THE COURT: Great. I am all for saving time.

MS. NATAPOFF: Thank you. Dr. Cole.

Whereupon,

SPURGEON COLE

was called as a witness for the Defendant, having been first duly sworn, was examined and testified as follows:

THE CLERK: Would you please state your name for the record, sir, and spell your last name.

THE WITNESS: It's Spurgeon Cole. You would probably rather me spell the first name.

THE CLERK: Yes, please.

THE WITNESS: S-p-u-r-g-e-o-n, and the last name is Cole, C-o-l-e.

THE CLERK: Thank you, sir.

DIRECT EXAMINATION

BY MS. NATAPOFF:

Q Dr. Cole, good afternoon. Thank you for being here.

A Sure. Happy to be here.

Q Judge Grimm has indicated that he has access to your resume. If you could briefly tell the Court about your qualifications and in particular the work that you have done in this area in field sobriety testing over the past decade?

A Well, I have in terms of degrees a bachelor of science, a master of science, and -- a bachelor of arts I mean, a master of science, and a Ph.D. degree all in the area of clinical psychology. I taught at Clemson University for 28 years, and my specialty has been in testing measurement and evaluation of tests.

In the last 10 or 15 years I've been particularly interested in field sobriety testing, basically the three tests that NHTSA has proposed, and I think the Judge has probably read the same thing -- many of the same things I have in NHTSA's '77 and the '81 and the '83, and there's three studies that have been in the '90s.

He may have read the article that I have written, and I have written several articles and --

THE COURT: And you wrote one for the South Carolina Bar Association, and you wrote one that was published in Perceptual and Motor Skills in 1994.

THE WITNESS: That's correct, and I've written a couple of articles for the Champion Journal, which was a critique of these studies. So I've written rather extensively in the area.

BY MS. NATAPOFF:

Q Dr. Cole, have you also testified on this matter in other Court proceedings?

A Yes. I have in several states.

Q Okay. Can you give us an estimate about how many times?

A In this type of hearing probably eight or 10 times.

Q And what is the time period that you testified in the past? A year, a decade?

A In the last 10 years.

Q Okay. I am going to ask --

MS. NATAPOFF: Just so you know where we are going, Your Honor, I am going to ask Dr. Cole to give us some of his general conclusions over his decade of research and expertise in this matter, and then I am going to ask him to go through the studies one by one.

BY MS. NATAPOFF:

Q Dr. Cole, if you could give us a little bit of a background. As a testing expert we are learning that you look at issues of reliability and validity. Can you explain what those concepts mean in the context of field sobriety testing?

A Yes. In evaluating a test in terms of its effectiveness and in terms of whether it's valid or reliable and the standardization there's certain procedures and techniques and there's certain ways that we need to try to address with the Court and explain to the Court exactly what reliability and validity mean.

THE COURT: Reliability means if I do the test with the same input I get the same output within an acceptable range of error. So if I jump on the scales and it is 100 pounds --

THE WITNESS: That's correct. In other words if --

THE COURT: And validity means we are measuring what is at issue in this case.

THE WITNESS: It simply means consistency. For example, if I give you a test today and I gave you the same test two weeks ago I should get the same score if it's totally reliable.

THE COURT: Right.

THE WITNESS: And so we're -- but reliability simply means consistency.

THE COURT: Right.

THE WITNESS: Now validity is the ability to predict something, predict a certain criteria. They're fairly easy concepts if we don't get them confused with common vernacular of reliability and validity.

But reliability is simply consistency of scores, and you try to get that first in a test. Then after you establish that you've got reliability then you can try to establish whether you have validity or not.

One of the problems with all the studies that we have in NHTSA except for briefly in the '77 study, they do not give us any reliability data -- I mean, I'm sorry, any validity data on the studies. They give us some reliability studies, and then they give us percentages.

Percentages, when you talk -- when somebody says that they were 90 percent accurate or 92 percent accurate, or 86 percent or 84 percent accurate it is really kind of meaningless.

Because when you give accuracy rates it's the lowest kind of report that you can give statistically, and it can be influenced by so many factors that it's not scientifically acceptable to report data in terms of percentages.

You have to report data in terms of reliability, validity, standard error of measurement, and standard error of estimate. That is the only way that you can tell or determine how well the test is doing because there's so many factors that can be -- that can influence percentages.

The first two tests, the '77 and the '81 tests, were done well. These tests were -- they were formed in the

laboratory. They used good controls. They used a double-blind study. They I think -- I have no argument whatsoever with the methodology that they used in those studies.

I do have some question about how they should have reported the data in terms of reliability and validity, but certainly they performed the tests correctly.

The three studies that they've done in the '90s, the Florida study and the Colorado study, are not -- would not meet the criteria for -- even minimal criteria for a scientific study, because in those particular cases they only give you -- first they don't have any controls.

They don't have -- they have confounded variables, and they have a base rate error, and they have numerous other difficulties with those. But the first two studies, the '77 and '81 study, we can look at that data and draw I think conclusions as to the reliability and validity of these three. The last three in the '90s are -- they're good for information, but they have no scientific value.

THE COURT: All right. Now, Ms. Natapoff, let me just -- we have been told they are ready for us next door, which means they are ready for me to walk in there and say "Start your presentation." None of you all have gotten any noontime refreshment, is that right? You are all sitting here with growling stomachs?

THE WITNESS: That's correct. We've had none.

THE COURT: Okay. Let's remedy that right now. Let's take what will be I hope no more than a 30-minute break, because once I get into this thing I am not stopping, and we are going to stay tonight until we get it done. And so let's let everybody take that break.

When we come back I have nothing else on the afternoon agenda, and, doctor, what you need to know is not only have I read the NHTSA articles. I have read your articles, I read your affidavits.

I have -- again I am glad you are not giving me that test for comprehension, but nonetheless I think I understand some of the basic points, and we can summarize it and give Mr. Marone a chance to ask whatever questions. Then we will go from there.

THE WITNESS: Yes, sir.

THE COURT: But I want to get the most for what we have to do here today. So let's feed your -- get your bellies so that they are not growling and interrupting your trains of thought. Make it 30 minutes.

We will try to be back here by quarter after, and I don't -- other than just brief comfort breaks we won't stop for the remainder of the day. Okay? All right. Thank you all.

THE CLERK: Please rise.

(Whereupon, a luncheon recess was taken.)

A F T E R N O O N S E S S I O N

THE COURT: Dr. Cole, could you come back up here please. You are still under oath. My apologies for the delay. I hope everybody had a chance to get a bite to eat and they are refreshed and rested. Let's go ahead.

Whereupon,

SPURGEON COLE

was recalled as a witness for the Defendant, having been previously duly sworn, was examined and testified further as follows:

MS. NATAPOFF: Thank you, Your Honor.

DIRECT EXAMINATION (Resumed.)

BY MS. NATAPOFF:

Q Dr. Cole, I can't swear to exactly where you were when we left off.

THE COURT: When we left off he said that the original studies in '77 and in '81 were good studies, but that the subsequent three studies from Colorado and Florida were not able to give the same type of information that the earlier ones had.

He was about to explain why it was that those studies were not scientific studies because there were no controls and they had multiple variables.

THE WITNESS: That's correct, Your Honor.

Finishing up on that statement, in the Colorado study and the

-- I think they did also one in San Diego as well as one in Florida. They never separated out the variables versus -- for example, they stopped people who were driving erratically, who smelled of alcohol.

They fumbled pocketbooks and things like that. In other words, they did not separate out the influence -- or the amount of information you were getting from the field sobriety tests from the information that you were getting from other sources.

So you cannot attribute the accuracy to what they found to the field sobriety tests because of so many confounding variables. They could have simply -- what they -- if they wanted to make it a scientific study they could have simply stopped half of the -- give half the subjects the field sobriety test and the other half the subjects not the field sobriety test, and then to see if there was a difference in the accuracy when you added the field sobriety test and when you didn't have the field sobriety test.

Then you would have a differentiation between -- then you would have known how much the field sobriety test actually contributed to the accuracy rate. The other, of course the primary problem with the study is that 79 percent of the people that they stopped were intoxicated.

Which means that you have a base rate here of 79 percent. So if you had been sitting in your station and

somebody simply says that this guy was driving down the road weaving back and forth, he smells of alcohol. Is he intoxicated? And if you said yes you would have been correct 90 percent of the time.

When I say this let me give another example in order to maybe clarify what I mean by base rate. Let's say I develop a test to predict suicide, and I come out with my data and I say my test is 99 percent accurate in predicting suicide in my clinical patients.

Well, by saying 99 percent it sounds like a good test, right? But since I know that there is less than one-tenth of one percent who are going to commit suicide if I simply say nobody is going to commit suicide I'm going to be correct 99 percent of the -- 99-plus percent of the time.

So percentages can be mislead -- misleading by the base rate. That's why you need to randomly select your subjects, get a base rate that is correct, so that you can now determine the amount of improvement that you get in terms of prediction from the tests that you use.

So not only did they confound the variables, they have a base rate of 79. Now they indicate I think that the accuracy rate was like 86 percent. Well, you have only -- when you have a base rate this high you are talking about only incremental improvement.

The incremental improvement was about seven percent

using the field sobriety test plus all the other tests that they had, the increment. So, you know, it doesn't give you any way of determining the influence or the effectiveness of the field sobriety test in determining whether the person was in fact intoxicated or not intoxicated.

Now, if you look back at the '71 -- I mean, I'm sorry, the '77 and '81 studies, they did all that. They had a random sample, or they had volunteers, which is about as close as you can get to a random sample as possible. They made it a double-blind study.

They calculated, you know, the reliability, and they didn't do validity, which they should have done, but they did at least give you a correct and incorrect -- number of correct and incorrect choices they made. So you can draw from that.

My criticism of '77 and '81 would be they should have given us the raw data so that we could have further analyzed the data to determine what we call the standard error measurement or reliability, and the standard measurement estimate in terms of how much -- how well they are able to estimate the BAC levels.

If we had that we could do a -- we could give you a little bit better understanding of what we mean by it, but we do have the data, and this is what -- this is the data, the '77 and '81, which I would like to go over with you and

explain the results and how good the results were and how -- I'm sorry.

THE COURT: I am going to let you hear -- I do want to hear that, but let me just ask the question, because I am afraid I will forget it if I don't ask now. What we are using the test to show is an important issue here.

THE WITNESS: Yes.

THE COURT: The Government in this hearing has conceded for purposes of this hearing and the Courts of Maryland have already said that the use of the NHTSA horizontal gaze nystagmus test is not admissible for purposes of showing the blood alcohol level of any given percentage of the Defendant.

So the Government will not be seeking to offer it for that purpose of direct evidence of intoxication. Now, the question that I have is there are any series of conclusions that can be drawn from various tests in increasing degrees of ambitiousness in terms of the conclusions they seek to draw.

One test might be that the consumption of alcohol can cause a condition defined as nystagmus, which is an involuntary muscular reaction to a stimulus, and therefore there is a cause and effect relationship -- a cause and effect relationship between ingestion of alcohol and the observable phenomenon of nystagmus.

Now that is level one. Assume with me right here. Level two is to then try to finesse that by saying that when you see this -- these three test results of this measure of this physiological phenomenon that you are now able to conclude to an error rate which is scientifically acceptable for test studies that this person has a blood alcohol level in excess -- greater than or excess of 0.01 percent.

THE WITNESS: I understand.

THE COURT: It is a different step. It is a much greater step than the first one, correct?

THE WITNESS: Correct.

THE COURT: Now, from the studies that we have seen about this and recognizing that just because one -- there may be one input condition that can cause another output condition, alcohol and nystagmus, and there could be many others, do folks such as yourself who have reviewed all the literature and critiqued the studies from NHTSA and others on this issue of a connection between alcohol and nystagmus recognize a legitimate causal connection between the presence of alcohol and the ability to observe nystagmus assuming you do not then try to quantify the percentage of intoxication or impairment by some percent based upon that?

THE WITNESS: Yes. There is a correlation between the alcohol and nystagmus. There is also a correlation of course between the consumption of alcohol and any psycho-

motor skills.

THE COURT: Absolutely.

THE WITNESS: The more you drink the worse you are going to perform.

THE COURT: Well, and they were doing -- they were convicting people of drunk driving long before they had the one-leg-stand and the walk-and-turn and the horizontal nystagmus test.

THE WITNESS: Yes.

THE COURT: Just by erratic driving, the classic "They smelled this way."

THE WITNESS: That's correct.

THE COURT: They were pugnacious in their attitude, they were slovenly in appearance, they were disoriented, they were slow, they fumbled with their keys. All those things, which have never been tested for anything.

THE WITNESS: That's correct.

THE COURT: And could be probative of lack of sleep or mental illness or depression or something like that. Not necessary, but are probative circumstantial evidence from which with everything else a fact finder might be asked to infer the presence of a condition.

But not, if you will pardon my expression, diagnostic of --

THE WITNESS: That's correct.

THE COURT: -- as in a sample of blood. Do you agree with that?

THE WITNESS: I would agree with that.

THE COURT: All right, sir. Now, I interrupted you. I apologize. I asked my question.

THE WITNESS: No problem.

THE COURT: Now I can forget it.

THE WITNESS: No problem.

BY MS. NATAPOFF:

Q Actually if I might just follow up on that question, because I think it is an opportunity to get into some other aspects of your analysis. I would like to take Judge Grimm's question one step further.

Which is having acknowledged that there is a theoretical relationship between alcohol and nystagmus, indeed alcohol and the deterioration of physiological function, could you discuss for a moment essentially the design of the field sobriety tests done by NHTSA in terms of their ability to measure physiological impairment versus the asserted criterion of the test?

Which of course is not measuring impairment at all, but measuring blood alcohol content, which is -- can be a substitute for impairment, but not necessarily a clear one. Could you explain that?

A There is a connection, a very poor connection. I mean,

the reliability of these tests in predicting BAC and the validity of these tests in predicting BAC, they are very poor. There is obviously some correlation that they have been able to establish between say walk-and-turn and the BAC levels.

The problem that you have is that this connection or this correlation is very low. So it is you have -- as the Judge indicates there is some relationship. The more you drink obviously the more you are going to get impaired.

But the field sobriety test, they do a very poor job of being -- first, they are not very reliable in terms of consistency. They are not very valid. They make numerous errors, but there is a degree of correlation between the field sobriety test and BAC level.

Q Now, are you using BAC level synonymously with impairment?

A No. There has never been a single study done in which they tried to relate field sobriety tests to impairment. Now, impairment in terms of driving impairment --

THE COURT: Actually the mechanical ability to function safely behind the wheel of a car. Is that what you are talking about, doctor?

THE WITNESS: That's correct.

THE COURT: Because I --

THE WITNESS: There has never been a single study

that shows any relationship between the field sobriety test and impairment. What you do is kind of a leap of faith that obviously we know that if you drink enough you're going to be too impaired to drive an automobile.

But they could do -- and I don't understand why it hasn't been done. You could actually do a study where you determine if the field sobriety test could predict driving impairment, which is the ultimate question that we would like to be able to answer. But so far there has never been a single study in this area.

THE COURT: I guess the thought is, doctor -- I do agree with you. I had an opportunity to try a case last year involving testing of steering rates on drivers which involved another NHTSA study, and they were able to test or attempt to test how reaction time varied with sudden emergency situations during various driving conditions and what people did.

One could imagine even if you did a driving simulator like an air traffic flight simulator that you might be able to do some type of test. But we use it as a proxy. I guess the theory is that if you drop your keys, and you forget who you are and you can't remember what exit you wanted to get off on and all of that, that there is a demonstration of concentration and coordination and reaction time that permits the drawing of an inference.

Therefore it is imperfect, but legislatures do that all the time when they say if you have got a point 0.08 or above we don't care what you can do. You can be superman and --

THE WITNESS: I agree, because we could -- you know, we call it per se law down in our state, and I agree. I mean, it's kind of an arbitrary decision to make 0.10 or 0.8, but I agree, and I think that's the way it should be done.

Because if you have -- because we can measure BAC levels accurately with proper equipment. What I guess our point is, the point here is that field sobriety tests, though, do not do a very good job of predicting BAC levels, or they do an extremely poor job of accurately predicting --

THE COURT: BAC meaning the percentage of blood alcohol content.

THE WITNESS: Percentage of alcohol in the system, yes.

THE COURT: Okay. Okay. Now, let me again -- I am going to shut up, because I promised I wasn't going to do this.

THE WITNESS: No, no. You go ahead. It makes my job easier.

THE COURT: No, no. Well, there is an old saying that the lawyers say. "Judge, if you are going to try my

case don't lose it for me." So I will shut up and let the counsel try their own cases for a while, but I do have some questions.

I have spent off time with all the information everybody has provided. I couldn't be more proud of the work that the lawyers have done. It is a tough area. The decision has got to be made by the person probably least qualified in the room to make a decision, but fortunately I am not shy. So go ahead, Ms. Natapoff.

MS. NATAPOFF: Well, Your Honor, I would kick it back to Dr. Cole.

BY MS. NATAPOFF:

Q I think you were going to embark on some specifics with respect to the studies and what they do and don't show?

A Yes. I was going to initially -- the first thing that you look for in a test is reliability.

THE COURT: Right.

THE WITNESS: And I was going to look -- it's on Table 14, page 35 of the '81 study.

BY MS. NATAPOFF:

Q That is about --

A I'm not sure that I'm going in the order that they were presented to you, but it's --

THE COURT: I got it. I have the study right here. Page what?

THE WITNESS: Page 35.

THE COURT: Got it.

THE WITNESS: Table 14. If you look here on the lefthand side, this is a study of test -- retest reliability, and --

THE COURT: Got it.

THE WITNESS: And the way they did this of course, and I think Your Honor knows, that you test the person and then you dose them to a certain level. Then you bring them back two weeks later and you dose them to the very same level and you retest them.

The question, if the test is perfectly reliable of course you would get 100 percent consistency. If it was totally random then you would get a zero. If you look at the nystagmus test, when you have the same officer -- this is the same officer on Subject A.

He gives the test to this person, Subject A, and then two weeks later they dose Subject A to the same level and they have the same officer give the field sobriety test to that individual.

THE COURT: Is this the 0.66 versus the 0.55?

THE WITNESS: Yes. Yes. Right. Okay. That 0.66 represents the reliability. In other words, that's how reliable that you can count -- that's how much reliability or how much faith that you could have in the score that that

individual gave the first time he did it.

THE COURT: Okay. Now, hold on one second, because you are spending somebody who has spent his entire life avoiding taking a statistics class.

THE WITNESS: Okay.

THE COURT: But when folks are designing tests for error rates that are considered acceptable, okay?

THE WITNESS: That's correct.

THE COURT: Is there within the business of psychological testing or human mechanical testing like this an accepted error rate for this reliability test/retest phenomenon that you would say?

THE WITNESS: Yes. The standard would be about 90 percent.

THE COURT: All right. So is that 0.9?

THE WITNESS: 0.9.

THE COURT: Okay.

THE WITNESS: Because 0.90 means that 90 percent of your score is accurate score and 10 percent is error. On the nystagmus test what this means is that 34 percent of your score is error; 34 percent of the score is error, 66 percent of the score is true variance, which is extremely -- obviously that is extremely low reliability for any type of test.

If you look at walk-and-turn, which is the next

one, and I won't go through all of them, but this indicates that there is 27 percent error in that score and a true variance of 72 percent.

THE COURT: And on the retest you could do them better flipping a coin is what you are saying, right?

THE WITNESS: Well, no. If you look across though, if you look at different -- there are two ways of doing it. If you look across now you are going to find the numbers go down. For example on the walk-and-turn when the same officer tests the same person at the same BAC level you get a reliability of 0.72.

If you look across, though, to a different officer, this is where you have a different officer testing Subject A at the same BAC level two weeks later. In other words, you have two different officers testing the same person.

Now the reliability goes down to 34 percent, meaning these two officers agreed -- only 34 percent of their score is valid when you have two different officers testing the same person at the same BAC level at a two-week period. So that is again -- and of course the one-leg-stand is a 0.61. It drops to 57 when you have a different officer.

If you look at nystagmus --

THE COURT: I thought that was .60. Did I miss something? 0.61 drops to 0.60?

THE WITNESS: Yes. Uh-huh.

THE COURT: Did I --

THE WITNESS: Which was one-leg-stand. Obviously there was 29 percent error -- I mean 39 percent the first time, 40 percent error when you have different officers.

THE COURT: Right.

THE WITNESS: Based on this, it would appear that the nystagmus test -- well, the walk-and-turn has the highest reliability of any of the tests. It has a 0.72. The nystagmus has a 0.66. The one-leg-stand has the least reliability.

Then if you drop down to arrest decisions that means that 56 percent of the score is error when you have the single officer making the same decision with the same person. Two weeks later he obviously makes the different decision. The bottom half of the page is observers.

I'm not sure we want to go into observers. I mean, they are not officers. They were just trained observers. Their validity is -- I mean reliability, I'm sorry. Their reliability is somewhere about the same, lower. Nystagmus was 55 with observers, 39 with walk-and-turn.

What this tells us is that most of the score that you get when you do a field sobriety test, most of your score or a large portion of your score is simply error. In other words, they can't measure it correctly. We are not talking about predicting anything.

We are simply -- they can't even measure, or they don't get the same scores when they test the -- give the same test to the same person at different times at the same BAC level.

THE COURT: Okay.

THE WITNESS: Okay. The next page I wanted to go to if I marked them down -- if I find my -- you get so prepared, and then you get up here and you can never find what you're looking for.

THE COURT: Well, it happens to every single one of us.

THE WITNESS: Oh, here I am. Okay. Next is 32, page 32.

THE COURT: Same study, sir?

THE WITNESS: Yes, 32, the same study.

THE COURT: If anyone who is following along can't find it holler before we start the explanation. All right. Go ahead, sir.

THE WITNESS: Okay. These are simply interrelated reliabilities again, and this is when you have two different people doing the same study on the same day. Interrelated reliability for nystagmus is 0.62, for walk-and-turn 0.74, and one-leg-stand 0.70, total score 0.78.

Again, we're actually talking about an absolute turn. In other words, if something is -- if it correlates,

if the correlation is 0.62 that means that there is 38 percent error, 62 percent true. You simply -- and reliability is -- they are not as complicated as validity scores.

For example, if it was 0.90 it is 10 percent error, 10 percent valid. These are --

THE COURT: 90 percent valid.

THE WITNESS: These are different individuals testing the same person at the same BAC level. Then I want to go to page 17, the '77 study.

THE COURT: Page 17?

THE WITNESS: Page 17, the '77 study.

THE COURT: Got it.

THE WITNESS: Okay. If you look at the bottom, the scores, the scores here. Now, these are not reliability scores. These are validity scores. In other words, this is ability to use the field sobriety test and predict a criterion or predict the BAC level in this particular case in this particular study.

Now, so you get a -- on the one-leg-stand you get a validity of 0.48, a finger-to-nose of 0.42, walk-and-turn 0.54, and they've gotten -- this is '77 when they went through about 10 different tests. Now, the one-leg-stand, that 48, what does a validity coefficient of 48 actually mean in terms of trying to understand it?

They don't give us a standard error of estimate, but I can tell you this. A 48 means -- this is what it means. It means that they are about 12 to 14 percent better than chance.

In other words, if you had simply let the person randomly guess the individual's BAC level and used the one-leg-stand you would be accurate about 12 to 14 percent more with the test than without the test. You have improved your accuracy rate about 12 to 15 percent above a guess rate.

THE COURT: I gather that you are going to tell me that that is not acceptable. What is the acceptable validity rate?

THE WITNESS: Well, acceptable, Your Honor, has to do with what you are using it for. If you are going -- this would be acceptable if you were going to hire individuals to work in a manufacturing company and you had 100 people who were applying and you only needed to hire one or two.

That would be acceptable because you don't have to worry about mistakes. If you are talking about, you know, maybe a jail sentence or going to jail you might want to have a higher validity than that because of the consequences.

I mean, that is not a particularly low validity for some types of decision making, but if you're making a critical decision, for example if you were going to place a student in a special education class or something like that

based on this score you certainly would want a validity much higher than that if you're making a crucial decision.

It did turn out that the nystagmus was the better test. It -- in '77. It increased the accuracy about 26 percent better than chance. In other words, if you randomly guessed -- had half your people randomly guess and the people use -- the other half of the people use the nystagmus you would do about 26 percent better than chance.

Okay. Page 25.

THE COURT: The same study, sir?

THE WITNESS: '77.

THE COURT: Yes.

THE WITNESS: I think this is the biggest problem with the test. If you look at the number of people in the column there under "officers" and under "arrests" and you look at the bottom figure you see 101 there.

THE COURT: Yes, sir.

THE WITNESS: Okay. That means that they arrested 101 people. Of the 101 people that they arrested, 47 percent of them did not have the BAC at BAC levels below 10. In other words, 47 percent of the --

THE COURT: Did not have it above 10.

THE WITNESS: Had below 10.

THE COURT: Right. They had below 10. They did not have above 10.

THE WITNESS: That's right.

THE COURT: Okay.

THE WITNESS: So in a sense you had -- they call them false alarms. In other words, the officers said these 47 percent of the people or 47 percent of the people that they arrested literally based on these tests should not have been arrested, and if you will read in the first paragraph here --

THE COURT: If the arrest criteria was 0.10 or greater.

THE WITNESS: 0.10. That's correct.

THE COURT: Okay. If the arrest criteria was whether or not had consumed alcohol and therefore may have been impaired along with everything else then it is a different issue.

THE WITNESS: That's correct. For example, if they had been at 0.8, but the officer was told --

THE COURT: No, no, no. I understand.

THE WITNESS: Okay.

THE COURT: You are saying under the way the study was done --

THE WITNESS: That's correct.

THE COURT: -- this is what they were told their go/no-go decisions were based on.

THE WITNESS: That's right.

THE COURT: Okay.

THE WITNESS: The researchers in the first -- in the paragraph says an error rate of 47 percent in making arrests is not acceptable.

THE COURT: Right.

THE WITNESS: They agreed that is not acceptable. That is too high. Then -- okay. Go -- I would like for you to look at number 29, the same study, '77, on number 29. This is the way they gave the reliability in the '77 study. They didn't give the numbers. They simply gave the reliability in terms of a scattergram here.

If a test is reliable all of those dots should fall along that diagonal. What you find here is that there is -- it's just a tremendous scatter. For example, under false alarms you've got one guy who scores a 60. If you're the highest one up there, he scored 60 although his BAC looked to be about 0.9.

Then you've got some guys down here with a 0.16 it looks like, and they passed the test, and this shows you here the -- if it had been reliable it would have been right along the diagonal. It looks more like a shotgun than it does -- approach than it does where there is actually any relationship between the two.

That's all the data they give us. They don't give us the reliability coefficient. They simply give us a

scattergram, and you can look at it and see here that the reliability is certainly quite questionable. We're only talking about reliability at that point, not validity.

THE COURT: And this is under lab conditions, right?

THE WITNESS: I'm sorry?

THE COURT: This is under lab conditions?

THE WITNESS: Yes. This is under lab conditions in '77. Under lab conditions in '77. I would like to refer you to -- now to the '81 study, page 22.

THE COURT: Page 22?

THE WITNESS: Yes. Before I did that, it might be easier to go -- let me -- if you don't mind, let's -- let me go to the '81, page 27 first. We can go back to that one if you don't mind.

THE COURT: All right.

THE WITNESS: I think as you recall in the '77 study they said a 47 percent rate was unacceptable. They went back in '81 and did the study again. This time they come up with a false positive rate of 32 percent. That's down from 47 down to 32.

Now, which looks like that you've made a significant improvement in the test, or you've made -- or you've better trained your officers. But what they did, instead of actually improving the test results -- improving

the test or improving the officers' training, what they did was change the distribution in order to get a better -- or fewer false positives.

If you will look on page -- the '77 study, page 12. If you will notice there -- excuse me one second while I pull this. If you'll notice that on the page 12 here they had groups of -- they had 72 people that had no BAC at all. In other words, 72 people.

Then they had 23 percent at a 0.05, then they had five percent of people at a 0.07, and then they had 21 percent at 0.10. Okay? Now, obviously if you are going to make errors the errors are going to occur somewhere between 0.07 and 0.10.

This is where -- because this is the critical -- obviously you've got 33 percent down here that's had no alcohol in their system, so you should -- the field sobriety test should be able to get those correct. Then of course the people -- you've got 17 percent above 0.15. You should get those correct.

Now, so that was the '77 study, and that's when they obtained 47 percent false alarms or false positives. So if you then go to page 19 in the '81 study -- I'm sorry. I'm sorry. Page 12 in the '81 study. Thank you. Page 12 in the '81 study.

THE COURT: Got it.

THE WITNESS: Okay. Now, if you notice, they dropped out. What they did is they now have -- they now put -- they only had 238 subjects in the '77 study. They had 296 in the '81 study. If you notice here now, they still put --

MS. NATAPOFF: Can I interrupt? Are we looking at the same --

THE COURT: I think we are on the wrong table. It can't be page 12.

MS. NATAPOFF: No. It is not page 12 in the '88 study.

THE WITNESS: It's page 12 in the --

THE COURT: '81. He said the '81 study.

THE WITNESS: '81.

THE COURT: There is no '88 study.

MS. NATAPOFF: I am sorry, in the '81 study, but page 12? Does that seem to -- the correct table to you?

THE COURT: That has got the LAPD and all those people.

THE WITNESS: Okay.

MR. MARONE: The background of officers?

MS. NATAPOFF: Yes. I don't think that is the one that you mean.

THE COURT: It is the wrong chart. What does the chart say on it, doctor?

THE WITNESS: This is the chart that I had.

THE COURT: Just show -- flash it so I can see it.

MS. NATAPOFF: That is still the -- that is also '77. Those are both from '77.

MR. MARONE: Try page 22 of '81.

THE WITNESS: I am sorry?

MS. NATAPOFF: 22 of '81.

THE WITNESS: 22 of '81? Okay. I'm sorry.

THE COURT: No. It is not that one, either. He is looking at -- he is still looking at that thing that looks like a -- looks like one of those --

MS. NATAPOFF: The '81 study doesn't have a chart like that, the information.

THE WITNESS: Well, it's in the '81 study. Whether --

MS. NATAPOFF: Yes. It just doesn't look like that. It looks -- it is on page 22. It is tables like that.

THE WITNESS: Okay. If you don't mind, would you mind continuing to look through that and see if you can find it in the '81 study? I would like -- I would prefer to have --

MS. NATAPOFF: But they don't do it in the '81 study in that form.

THE WITNESS: Would you mind double checking, double checking for me?

(Pause.)

THE WITNESS: Page 15 would be -- I'm sorry. That's the page we're looking for.

THE COURT: Page 15?

THE WITNESS: 15 on the '81 study. Yes.

THE COURT: The one that says figure four, drinking history?

MS. NATAPOFF: Yes.

THE WITNESS: Page four, drinking history. Yes.

THE COURT: Okay. Got it.

THE WITNESS: Okay. Now, if you compare -- I'm sorry to have confused everybody. If you look at this particular --

THE COURT: Hold on a -- are we still on the record?

THE CLERK: We are on the record, but I don't have the mike on.

THE COURT: Okay. That is all right. Just talk loud, doctor. I will hear you.

THE WITNESS: I can do that. I can talk loud. If you notice in this particular -- in the '81 study they continue to put 97 subjects or 33 percent in the zero alcohol column, but instead of having a point -- then they had 34 percent of the people in the 0.05 column.

They had nobody in the 0.07 column, and then they put 22 percent in the 0.11 column. Now, what that means is

that what they did from '77 to '81, they decreased the number of subjects that were in the difficult zone. You actually went from like a 22 percent in the difficult zone in one to a 34 percent difficult zone in the other.

So the reason for the improvement from the -- and I don't -- maybe this going way more than we need to do, but the reason they went from 47 to 32 error was the fact that they spread the distribution out, had more people, didn't have as many people in that center area in which the officer would make mistakes.

That I think accounts for the improved score, not improved in testing or the evaluation of the test or training of the officers. Okay. Did we do page 27?

THE COURT: We did 27 of the '81 study.

THE WITNESS: Okay. Now we're doing -- okay. Now --

THE COURT: And 22 I thought was the next one you wanted to do.

THE WITNESS: Then 22 was -- okay. Then number 19, study -- table 19. Well, actually we haven't done 22. Let's do 22 first if you don't mind.

THE COURT: Page 22 of the '81?

THE WITNESS: The '81 study. Yes.

THE COURT: Got it.

THE WITNESS: Okay. This table here, I put this in

or asked to talk about it in order to show --

THE COURT: Mr. Marone, if you can't hear position yourself so you can. Can you hear him all right?

MR. MARONE: I can hear him fine, Judge.

THE WITNESS: Can you hear me fine? I can talk even more loudly.

MR. MARONE: I can hear you fine, Dr. Cole.

THE COURT: See if it does work, Joyce, if you could.

THE WITNESS: Maybe I can talk --

THE COURT: Okay. Keep trying.

THE WITNESS: Maybe I can move a little further away from it or --

THE COURT: That is fine where you are, doctor. Go ahead.

THE WITNESS: Okay. If you look under placebo dose, they rated 18 percent, almost one out of every five individuals, the officers did, 18 percent of the individuals who had zero BAC tested by them, they said 18 percent were impaired.

For people who had BAC levels of 0.05 they indicated that 31 percent of those individuals were impaired. That is a high level of error for making decisions of impairment when you're talking about one out of every five individuals who have no BAC level who is actually considered

to be impaired and 31 percent of the individuals with a 0.05 were considered to be impaired, which again suggests that you have a very high false alarm rate using these particular tests.

There was a great deal said about nystagmus this morning, and I wanted to cover page 31, table 11.

THE COURT: The '81 study, sir?

THE WITNESS: Yes. I think it's the '81 study.

THE COURT: Page 31?

THE WITNESS: Yes.

THE COURT: Yep. Table 11.

THE WITNESS: Yes. This a correlation between the machine angle nystagmus onset and the individual rater's estimate of onset. In other words, this is a comparison for how well the officer, a trained officer in a lab situation can estimate the angle of onset versus the machine.

As you notice here, observer one, 0.34. That means about 65 percent of his score was error. Observer two, 53 percent of his score was error. As you look down the line here you can see the tremendous amount of error, but I don't know what officer number four was doing, but he has 0.23, which means that --

THE COURT: He would be better flipping a coin, right?

THE WITNESS: He's 77 percent wrong, and I don't

know where he was during the training period. Then officer five was at 26. It is interesting that officer one was at 0.72, but of course I guess out of 10 somebody is going to do well and somebody is going to do poorly.

If you look at the overall results the correlation is 0.58. This suggests certainly that officers cannot estimate the angle of onset of a nystagmus very well. They do a very poor job of it when compared with a machine.

THE COURT: And the methodology was changed in the second go-around. The first one that had that little gizmo set up with a chin rest, right?

THE WITNESS: This is the chin gizmo, yeah. I think -- yeah. That's right. They used the chin gizmo, yeah, I think.

THE COURT: Right.

THE WITNESS: And then they used the machine to do it. So obviously if we assume that the machine is correct these were the correlations between these individuals and the machine.

THE COURT: I see.

THE WITNESS: Suggesting that obviously you have a very difficult time with estimating the angle of onset.

THE COURT: Got it.

THE WITNESS: Okay. Those are the basic elements of the two tests. I could talk more about the other three

tests from the others, but I don't think there's hardly any value in doing so. I think you've read those and you understand the limitations of those studies.

I wanted to point out here the error rate and exactly what the reliability and validity suggests and this sort of things, and I'd be happy to answer questions from any of the attorneys or yourself, either one.

THE COURT: Now, was there -- does it make sense -- are you going to go on to another area after this then?

THE WITNESS: Well, I was going to talk just a moment about the '83 study, but --

THE COURT: Okay. Okay. Does it make sense -- I am not telling you, I am just -- does it make sense, Mr. Marone, if you have any questions on where he has been so far to ask him now and then -- so you don't have to try to come back and -- or do you want to wait until the very end and ask questions?

MR. MARONE: Your Honor, I don't want to intervene in the flow of the Defendant's case, but I just had some questions on some of the charts. I almost feel like we are here in almost like a briefing or in like an academic setting where it would be nice to raise your hand and say, "I got a question about that." If I could do that, I would prefer it.

MS. NATAPOFF: I have no problem with that.

THE COURT: Ms. Natapoff says no problem, and we

are seeking -- this is why it is a preliminary hearing and we can think out of the box. So that is what we will do. Go ahead and ask your questions about the charts, sir.

MR. MARONE: Okay. Thanks.

CROSS-EXAMINATION

BY MR. MARONE:

Q Dr. Cole, I just had some questions about some of the points that you made.

A Well, I am an academician, so maybe that's why it sounds like that. Sorry.

Q And we appreciate it.

THE COURT: So we have a low error rate in guessing what your background is.

BY MR. MARONE:

Q Sir, you made some comments about that in the original 1977 study there was some spread over different BAC levels, including a 0.07 BAC level, and then in the 1981 study they eliminated the 0.07 criteria and then just kind of left it from 0.05 to 0.10.

Now, do you -- are you just speculating that that is what they did, or were you involved in that study, you know that they changed that criteria to manipulate the results?

A I'm just saying they changed the criteria.

Q Okay.

A What they did, what they did is they actually dosed a

person to a 0.05, not 0.06, 0.07, but 0.05, and then they dosed people at 0.11 and then at 0.15. I mean, that's what they said they did.

Q Okay.

A And I'm not accusing them of nefarious behavior here. I'm just simply saying reliability and validity depends on the distribution. For example, if I developed a test that would predict success in college and if I gave half my subjects for people from a special education class and I picked half my subjects from the gifted class, my tests would look very good because I've got a distribution here that's going to give me the very best scores.

So I'm simply saying the distribution changed, and the distribution changed in favor of making more correct decisions.

Q Okay. And if I could just refer to your packet that we have been referring to. Page 17 of the 1977 study is one of the first documents that we looked at. It is the little figures at the bottom of the page under the heading "Do the tests discriminate impaired drivers?"

I just wanted to ask you, you said that the total nystagmus was 26 percent better than chance.

A That's correct.

Q And I just -- can you maybe just explain that concept to me. What is chance, and then what is 26 percent better than

chance?

A Chance would be just a random guess. I mean, for example if you didn't see -- if you didn't give any test, you didn't do anything. You simply asked the person to guess the BAC randomly, without seeing the person, without anything.

Just a random guess, and then if you have the person give for example the nystagmus test the person who gave the nystagmus test would be -- would do 26 percent better at predicting the BAC than the person who just made a random guess.

Q So would chance be basically 50/50? Would it be a 50 percent chance of getting it right? Would that be --

A No.

Q No?

A You are talking about predicting BAC level.

Q But, as far as --

A How close that you come to predicting the BAC level. See, when you are doing validity in this particular case you are not saying impaired or not impaired. It's not a yes or no. The officer is asked to predict the BAC, and he would predict the BAC 26 percent better by having had the nystagmus than he would if he didn't have anything.

Q Okay. But he because he is -- it is not just a choice of one or the other. It is multiple choices. There are multiple levels at which he could place the individual?

A He could have picked any number between zero --

THE COURT: Zero to 0.3, legally dead.

THE WITNESS: Yes. I think it's 0.5, but --

THE COURT: Is that right? I haven't seen that yet.

THE WITNESS: But he could have gone to 0.5, but I'm assuming he didn't predict that hard -- high.

BY MR. MARONE:

Q So the officer then is -- just by random chance the officer picking one of these, either 0.05, 0.07, 0.10, whatever, there would be just a random chance that he could pick that regardless of using a test, and then by using this particular test he betters his shot by 26 percent? Is that fair to say?

A That's correct.

Q Okay. And then with the -- you talked about the one-leg-stand test. He betters his shot at getting it right 12 to 14 percent?

A About 14 percent. Yes.

Q What about the walk-and-turn test?

A He's about 16 percent better.

Q 16 percent?

A Yes.

Q Sir, the two tests, the '77 and the '81 test, were based on a BAC level, an intoxication level of 0.10?

A That's correct.

Q And so the -- what they were all shooting for was trying to identify when somebody was at the 0.10 level and when they weren't?

A That's correct. Above or below.

Q Above or below. And since that time, times have changed, and is it fair to say officers have become more acclimated to providing field sobriety tests since 1977 and 1981? It is more of a common occurrence in law enforcement now?

A I would assume so, although these people were specifically trained by NHTSA here. So I would assume they were highly trained officers. They only had 10, used 10 officers, and they trained them over a several-day period and had them go over and do cases and things like this. So they were well trained individuals.

Q And they were well trained on detecting intoxication which would equal BAC at 0.10?

A That was their criterion point.

Q That was their ---?

A That was their criterion point.

Q And now times have changed, and now we are looking for folks that are impaired or under the influence of alcohol at a lower BAC level, 0.08 in the State of Maryland now, and we assimilate. We borrow Maryland State provision in our

federal jurisdiction within the State of Maryland.

The fact that has changed, does that make the test -- does that basically throw out the '77 and '81 study as no longer pertinent to the present situation that we are confronting now? Do we just throw it out?

A No. I don't think -- I wouldn't think you would throw it out. No.

Q Okay. Does it make the test because there were a lot of people that were determined to be false alarms back in '77 and '81 under the -- said to be over the intoxication level of below the 0.10, now they would actually be under the influence or impaired under the present state of the law.

Would they then now, you know -- would that meet the test of reliable indicator of impairment?

A No. I don't think you can look at it that way. It's kind of like if you raise the speed limit to 120 miles an hour you don't catch many people speeding. So the criterion that they were asked to use was 0.01 (*sic*). If you did the study again and you ask 0.08 obviously they would lower their criterion, too. So I don't think that would make a difference.

Q The tests that were used in '77 and '81 and the tests they determine that should be used in the '77 study that are now, you know, part of the standardized field sobriety battery, are they a good indicator of impairment if they are

done correctly by a trained officer?

A Well, again, you are going to -- let's talk about -- instead of talking about impairment let's talk about BAC, because, I mean, there is no connection here. We don't know of any connection between impairment, if you mean impairment in terms of driving. So --

Q See, unfortunately we -- you know, talking about BAC isn't very helpful to us because we are not going to -- you know, we are not going to say that based on your performance on these three tests it is more likely than not that you have BAC level of over 0.10 or it is more likely than not you have a BAC level of 0.08.

It is really just kind of a piece of the puzzle that goes into determining whether or not you are under the influence of alcohol or you are impaired. So I guess we want to see about how reliable and valid are these tests in making that determination. I mean, can we --

THE COURT: Let me just frame it this way, doctor, just so you know kind of the framework that we have to do -- while there is a tremendous amount of similarity I would guess from one state to another in terms of the basic structure of their DWI laws it is not a one-size-fits-all.

The Maryland Legislature at least for purposes of where we are in this particular case has changed recently. They went through a lot of different varieties of this thing

before they got to the one that was the most recent, and what they established was if you had a 0.01 it was -- 0.10, right? 0.10. Thank you. 0.10.

It was driving while intoxicated per se. Then there was driving while intoxicated, which means you had to find they were intoxicated, but there was no -- you sort of had to just guess by whatever queues you had.

They were driving 95 miles an hour in reverse going the wrong way on the highway throwing beer cans out of the front seat every 10 seconds, you know, and -- you know, singing drinking songs. That kind of thing, and you would say, "My gosh, they were intoxicated," meaning they were as bad or bad as you can get.

Impaired was, well, you are not intoxicated. The distinction of course is the punishment. If you are intoxicated it is a year and $1,000. If you are impaired it is $500 and two months. So the distinction is the level at which you are.

THE WITNESS: I see.

THE COURT: And if you are impaired you are not intoxicated, but we don't have a specific blood -- they tried all that. --- and it just didn't work, because what happened was with the blood test being optional you could refuse it. You know, the word gets out the way the word does, and a lot of folks who are -- they weren't so impaired that they took

the test, you know, when they were given the option.

So the words "impaired" and "intoxicated" are words that have legal significance.

THE WITNESS: Okay. I understand.

THE COURT: Intoxicated being the most serious possible level of intoxication, thank you very much, and there is a whole stack of cases that say what intoxicated is and what the difference is between intoxicated and impaired.

Now, they then added this driving while intoxicated per se to get this blood alcohol level so that for those that did take the test then we didn't have to get into to the debate about "Well, I'm a football player, and for me 0.10, I'm just getting started. I got to get up to be about 0.3 or 0.4 before I'm in any trouble because, you know, I've been -- this is the kind of metabolism I have."

You didn't have to get into all that nuance it created as a matter of law. We don't need to go with the effects this has on your actual ability to steer and your reaction time, and your judgement and your emotional mood, and all those other things that you and I and everyone in this room would say what makes a driver impaired.

What functions do they do less well, what are the critical functions to driving safely, and how are those adversely affected by the ingestion of alcohol? We never got there. The law just doesn't give us that term.

So when the officer comes in and says "impaired" you can do it by any one of a million different ways, and really that is -- I am putting this out on the table because that is the job we have got to try to figure out from the facts and data that are out there.

Many a time an officer came in and they had no walk-and-turn, they had no horizontal gaze, they had no standard field sobriety tests. But they would come in and they would say, "I was driving on the road. The guy passes me going 90 miles and hour. I go into pursuit. I put the sirens on. I'm driving for 45 minutes after him. He's changing lanes without signaling."

You know, "He's burning rubber. He's chasing people off the sides of the road. He backtracks and goes the wrong way. I finally get the guy to stop. He gets out. He's staggering. He can't find out -- he doesn't know who he is.

"He's 90 miles past where he thought he was going. He's going the opposite direction that he said he was going. He falls down. He's laughing. He's crying. He's all those things." And the jury said, "By golly, he's intoxicated," and no one ever -- I mean, in that case we had how the person looked when they were driving the car, and you could draw those inferences.

So I guess the real key is I think that the

evidence is probably -- and the Government said that. "I am not trying to offer these tests to show an ability to predict accurately by scientifically measurable, valid, reliable tests that can validly predict this, the ability to say the blood alcohol level is 0.10 versus 0.09 versus 0.07."

What he is going to try to say that the evidence allows and what Ms. Natapoff is saying it doesn't is that the -- well, someplace in between. You could have a lot of different things.

You could have the officer come in and say, "I'm not going to tell you what I think the blood alcohol level was, but I got nystagmus, and I know that physiological nystagmus can be caused by alcohol. So the fact that he displayed a physiological reaction that has a causal connection with alcohol and the fact that when he did the one-leg-stand he couldn't do it, and the fact that he counted only two steps forward and two steps back and didn't turn right and didn't pay attention to my instructions, and the fact that he smelled like a brewery and was confused and it took him a lot of time, then that supports the conclusion that he got that way by ingesting alcohol."

Now, the Defense could always come in there and say, "No. He's got seizures," or "He's got bad knees," or "He can't hear," or whatever else. But that is how you get facts from which the inference can be drawn, and that is

where we are trying to get at, is what use if at all can these tests be had.

THE WITNESS: I understand the situation. One of the problems that I have with like the one-leg-stand and the HGN and things like that, most states as you had indicated earlier have gone to saying you can give observations but you can't call them tests. You can't say pass or fail, you can't clues and things like this.

THE COURT: Right.

THE WITNESS: But I think what happens though is a juror is able to judge if a person is going down the road at 90 miles an hour going backwards, can't tell his name, and doing things like this.

The problem I think is, the greater problem is, how do they evaluate a one-leg-stand, because I don't think they have -- you know, they don't have the experience or the background to understand the one-leg-stand. So every individual in that jury knows the one-leg-stand is part of a field sobriety test.

So when the person says -- when the officer says, "He can't stand on one leg steady, he couldn't do the one-leg-stand," I'm afraid the assumption might be that they have flunked the field sobriety test but -- and not being able in their own mind to have the criteria.

Because obviously falling down is something we all

would know would be a very good characteristic. I'm not sure how they could judge the one-leg-stand or the heel-to-toe or something like that would be a concern that I would have.

THE COURT: The way we have to deal with that is in our discipline, doctor, is we have threshold tests of relevance where the Court has to first decide whether this fact has any tendency to make some other fact more probable or less probable than it would be without it.

So if we had no information at all except the person was falling down, would that have some tendency, and what I like to tell my evidence students is if they can't figure out a way to say it is reliable then they ought to be writing wills instead of trying cases.

So relevance is -- relevant. I said reliable. I meant relevant. So relevance is usually there. Then there is whether if it comes in is there an unfair danger that the jury is going to get it wrong, would be mislead, or have unfair prejudice, in which case the Judge can throw it out even though it is not relevant.

THE WITNESS: Yes. I see.

THE COURT: And the final way is you have a rule that says evidence can come in for one reason but not another, and we have what we call in the law as a fiction. That is gussied up to be a bald-faced lie.

But we have what we call fictions, which are legal

pretends where the Judge says to the jury, "Ladies and gentlemen of the jury, you have seen evidence that this person can stand. He tried to stand on one foot and could not do it when the officer asked him to.

"You are not to take this as being any indication as direct evidence of intoxication or any particular level of intoxication, but simply as one factor for you to evaluate if you find it helpful to get to the idea of whether they were impaired." So we try to deal with that by trusting juries to follow our instructions.

THE WITNESS: I understand that. I think that would be an excellent effort to try to make it --

THE COURT: And that is why you see the Court struggling to try and sort of get a middle ground here on these things frankly, and that is I think where the key is to this hearing that I have because the lawyers are so good and they have bracketed the issues as that tough area.

I mean, we are in -- for purposes of scientific reliability or judicial reliability, I am in that range where we are not at the 0.0001 or the 0.34. I am in that range you were talking about with the difference between 0.7 to 0.10 where I am hoping to do better than chance in my ruling.

MS. NATAPOFF: Judge, since we are apparently in class, if I could just add something maybe directing a couple of questions to Dr. Cole on the point that Mr. Marone raised.

THE COURT: Yes. At any time someone wants to stop this process and get back to the traditional way, fine, but I think we are getting information now, which is I think --

BY MS. NATAPOFF:

Q I would like to follow this path a little longer because I think it sort of -- it goes to the heart of one of the questions we are trying to answer, which is how -- what is the value of the activity that makes up the walk-and-turn or the one-leg-stand. What does it tell us, and if I could ask you to answer to separate questions.

You voiced a concern about what jurors will assume if they are told someone can't do the one-leg-stand, but you also -- we all acknowledge that if someone is walking down the street and they stumble and they fall, and they poke their finger in their eye, and they do other things that are generally considered to be indicia of intoxication we don't have any problem with regular people drawing inferences from that.

Are you saying -- is there any difference in the one-leg-stand or the walk-and-turn from those kind of normal --

A Well, certainly. First, we don't know what the norms are for standing on one leg. We don't for example what was the average -- how long can the average person stand on one leg? Is there a difference between a 20-year old and a 40-year old?

We don't have any -- you know, they never developed any norms to tell us what we should look for. That's the article I think you probably read. We wrote an article called -- I think it was labeled what -- designed failure to -- they are designed for failure.

Because they are not -- there are no norms. You don't know where the -- what a 20-year old versus 30-year old versus 40-year old versus 50. We know that they say don't give it over -- to somebody over 65 or somebody over --- pounds, but other than that we -- you know, you compare 20-year olds with 40-year olds, and I don't know. Judge, this is a pretty difficult issue.

Q So the one-leg-stand is not the same as watching a person walk down the street weaving. Is that fair to say?

A The way we did -- I did a study, and I'm not claiming that I did a study as thorough as NHTSA's here. But we did a study where we gave people what we considered normal tasks to do versus the abnormal, what we call abnormal tests.

Q And this is your article on perceptual and motor skills, right, 1994?

A Perceptual and motor skills. Yes, and so we asked them like, you know, their name and their social security number or their address and things like that. And we videotaped it, and then we gave them the field sobriety test, and we videotaped that.

And then we gave officers the field sobriety test and we gave officers observing them doing what we call normal things like just simply walking back and forth, giving phone numbers and addresses and things like that. We got the same results that NHTSA did.

And 47 percent of the time the officer said their individual was too drunk to drive when in fact none of our subjects had any alcohol in their system at all. Now, but when they looked at the person doing normal tasks there was a less error rate.

I mean, I don't recall, but I think it was like -- on about 21 or 22 percent of the people they considered to be intoxicated watching them do the -- observing them doing what we would consider normal tasks, and I'm not saying that's the way to go.

I'm simply saying that it might be better or one might consider to have them doing -- you know, taking out their pocketbooks and fumbling with their pocketbooks, fumbling with their drivers license, having them walk back and forth is -- you know, it would be -- Dr. Burns argued with me that this is not sensitive enough.

My argument with her is the standing -- one-leg-turn -- I mean heel-to-toe and one-leg-stand are too sensitive. That they -- you have a lot of false positives when you do it one way. She argues, well, if you do it my way you're going to

have a lot of false negatives.

So, you know, you're going to have a lot of people walking around that's driving that's going to be intoxicated because they can do these tests but can't do the other tests. So I have sympathy for the Judge in making the decision here because I know that the jury can't evaluate certain things like falling down and this sort of thing, and mumbling and slurring speech and bloodshot eyes.

I think they have more difficulty with the one-leg-stand because if we had some kind of norms we could tell them, if we had some kind of, you know, how much -- you know, how much value you could give to a person, what's the average -- how long can an average person stand on one leg, you know, who is 20-years of age.

I think maybe the jury could evaluate it better. My theory would be that the juror is not in a position to evaluate, you know, the scores on a one-leg-stand or a heel-to-toe or nystagmus and things like that.

Q Because that -- forgive me for putting words in your mouth. Tell me if this is incorrect, but that -- so what the NHTSA studies do and what training does for police officers, it gives them -- it gives those tests meaning.

It is specialized expertise, and although you disagree with the results of those studies and the reliability of those studies, isn't NHTSA's effort to infuse the one-leg-

stand and the walk-and-turn with meaning that a police officer can testify to? They say, "Well, you missed two clues. It's --"

A Well, I'm not arguing with their validity or reliability. I'm just simply saying what it was. You know? I'm not arguing it. I mean, I'm saying the studies were done satisfactory. Yes. I think obviously that's what they were trying, you know, with their effort is to -- as Dr. Burns had indicated to me, what they wanted to be able to do was take an officer that had been -- you know.

You know how defense attorneys attack an officer who has only been on the job one year versus -- he says, "Well, you've only been one year, right? And you -- so your evidence is not as credible as if we had an officer of 15 years."

So if the officer could say, "But I gave the field sobriety test and this is the scores." You know, "He passed this and failed this," or this something then this gives a first-year officer a -- more credibility. I mean, and her goal of course is like all our goals, to keep intoxicated people off the road.

I mean, that's as we said earlier. We -- it's a major problem in the United States. It has improved, but it's a major problem and it's a major decision to make in the cases.

Q I think if I might move the discussion along --

THE COURT: Whatever you want to do, Ms. Natapoff. Go ahead.

BY MS. NATAPOFF:

Q I think you were going to talk about the 1983 field study, and just as part of that I was going to ask you to talk generally about the difference between the lab and field studies in terms of what they can tell us about reliability and validity.

Obviously there are base rates in the field studies that don't exist in the lab studies, but if you could try to explain what lab studies can tell us that field studies can't?

A Well, lab studies you can control all the variables. You can control the temperature. You can control exactly what is said. You can control exactly the amount of alcohol. You can do a double-blind study. You can control all the variables, and the only variable that changes was the amount of alcohol in the subjects.

In a field study you don't get a random sample. In the '83 study they didn't test some people that looked okay when they observed them. They looked okay so they didn't do -- they did not score them in the '83 study. Most of them had the breathalyzer test right there before they gave the field sobriety test.

If I might point out one more -- okay, and then of

course they were precluded -- I mean, when you are out in the field your job is to enforce the law, and so you're not so concerned with running a test. But it is possible, and they had the opportunity to do this three times.

They had the opportunity to do the study right in Florida, they had the opportunity to do the study right in Colorado, and they had the opportunity to do it right in San Diego. And all three times they failed to use proper or even come close to using proper scientific methods to do this.

I mean, all they would've had to done in San Diego for example -- they had the BACs at the time that they gave the field sobriety tests. All they had to do was give -- have observers. Have the observer, make them fill out the form, estimate the BAC, and then have the observer walk up and give the BAC.

And then we would have known how well they could use -- how well the field sobriety test predicted that. But instead in San Diego they give them -- they don't have any observers. Then they go to Florida and they do the test. They have observers in two-thirds of the cases, but they don't use the BAC to determine.

You know, and I'm -- so the opportunity has been there to validate these tests in a field study, although admittedly it's difficult. You've got more variables that you have to control, but it's certainly possible to do.

Q I think one of the things that we struggle with in reading the studies, Dr. Burns herself I think is in the Colorado studies. She sort of distances herself from the lab study.

She says, "Well, that stuff was done 20 years ago when it was under lab conditions. We are not worried about what happens under lab conditions. We are worried about what happens in the field because this is where this stuff has to work, and in the field it works." Why is -- what is wrong with that?

A Well, it is the right opposite. It is going to work better in the lab than it is in the field. I mean, in the lab you've got everything going for you. You've got -- you've trained your own officers. You've trained your own subjects. You've got the BAC levels that you -- that are exactly right.

In a lab study the problem is always whether you can take the results from the lab and generalize to the field, not vice versa. If you can't do it in the lab you certainly are not going to do it in the field. I mean, you are going to get your best results in a lab study than you are in the field study.

Q And when you say "best results" you mean the most reliable?

A You're going to get the most reliable, most valid,

because you've got better training, you've got better -- You've got temperature. You've got the room. You've got fast services. You've got no flashing lights. You've got everything standardized.

So you should be able to get the better results there, and out in the field you've got different times of the night, guy gets stopped at 3:00 in the morning and the lights are flashing. He's fatigued. I mean, you don't have -- you have no control over that.

You don't have control over the age. You don't have a random sample. So you're going to get better results on lab than you do the field studies. If Dr. Burns thinks the field study would do -- would be better, I mean, they ought to conduct a field study to show it.

Q I didn't mean to cut you off, and if you were going to discuss the 1983 study a little more --

A I just want to go back to one table. Your Honor, if you would look at table 35. I'm sorry, page 35, table 14.

Q Is that the 1981?

A Yes.

THE COURT: All right. Let me make sure I got the right test here. Page 35? Got it.

THE WITNESS: Pardon me.

THE COURT: That is all right.

THE WITNESS: If you look down at the BAC, where it

says "BAC"?

THE COURT: Right.

THE WITNESS: 97. That's the reliability of the BAC, 97. I mean, and if you go down to -- if you look at even observers it's 96. You know, the BAC, there's no argument about that. That is a good reliability score.

THE COURT: That is with the breathalyzer or the blood sample?

THE WITNESS: Yes. In this case it was a breathalyzer. So, I mean, these are obvious. So you can see when you compare nystagmus and walk-and-turn and all that with the breathalyzer how much more poorly you're doing with these other tests than you are with the breathalyzer.

And I understand that sometimes you can't get breathalyzers, right, and you have to use lesser --

THE COURT: You can't get them if they don't consent.

THE WITNESS: Yes. In our state you -- I think you lose your license for a year if you don't take the test.

THE COURT: Well, they have administrative consequences which can result in the losing of the license.

THE WITNESS: But obviously -- I mean, I pointed that out to just simply show that that's what a good, reliable test would indicate. It would be -- a 97 would be good.

BY MS. NATAPOFF:

Q One of the things legally that we ask of tests is what is the error rate or is there a known error rate. There are five studies, two lab tests, three field studies. Do we know what the error rate is for field sobriety tests?

A No. We don't know the standard error of what we call standard error measurement or the standard error of estimate. No. We do not.

Q And that is not because you couldn't know. It is that simply the studies haven't been done to tell us what it is, is that correct?

A Well, I've had -- I was in Colorado Springs and they argued -- the prosecutor there argued that you can't just stop people randomly and do these tests. You can't just walk up to a person and give the breathalyzer test, and therefore the studies can't be done, and I -- and if they can't be done, they can't be done.

I don't know, but if you could stop subjects on a random basis and half of them give BAC and half of them don't then you can establish enough controls in order to get a valid estimate of what it would be or the validity of them in a field setting.

That may not be possible based on the laws. I don't know. I'm not a lawyer.

Q --- the three field studies that many of the

participating officers in those studies had portable breathalyzer tests available to them maybe even before they gave the field sobriety tests, but at least in the mix of the reporting. I thought maybe you thought about --

A It was kind of interesting. A matter of fact, if you look at that '83 study, I just thought of it, Maryland had a -- I think the accuracy rate was like 94 or 95. Of course they did have the BAC level at the time that they gave the case, but they were supposed to give the test and then look at the BAC.

Either Maryland is awful good, much better than North Carolina and other states, or they didn't have observers. They did not have observers with them and whether -- and, you know, there's pressure on the officer, too. I mean, if I was an officer and I got that, I got the BAC and I've done my test, I don't know.

You know, it's kind of like giving my students I guess the exam scores and then giving them the exam, and say, "But don't look at the scores until after you've taken the exam." I don't think you could do it that way. I mean, I think NHTSA says that you've got to have observers to be -- before you can draw conclusions to the thing.

Q As a matter of testing, sort of establishing the reliability of the field sobriety test, one of the legal issues is that the portable breathalyzer test is itself not

admissible evidence, whereas of course the intoxilyzer results that were used in the lab studies are.

But in the field tests in 1983 in Colorado and Florida presumably the correlation between the accuracy of the field sobriety test was being measured against the PBT result, which is itself not considered reliable and not considered admissible.

Is the PBT reliability, is that -- given the way the study works, is that going to be a cap on anything you can say about the reliability?

A It is. The PBT is a lot more reliable than nystagmus or the one-leg-stand or the heel-to-toe. I guarantee you that.

Q In what way? Can you expand on that a little bit?

A Well, I mean, it's going to be -- the reliability of the PB is going to be up around 90, 91, 92. It is not going to be as high I suppose as the one that you -- the big machine, but the PBS, I mean those machines are very good.

I mean, you get a reliability of up in the 90s which makes it difficult to understand how -- from a legal standpoint how those are those -- you know, from us laymen, how those are not acceptable for evidence and something else with much less reliability would be acceptable or admitted.

Of course, I understand they judge observations rather than scientific, and I think that's probably the difference.

BY MR. MARONE:

Q Sir, Dr. Cole, we were talking about all these tests and retest reliabilities. Does that provide us any information about an error rate at all? We don't even get to the point where we can even determine an error rate? Is that --?

A Well, you have to have the raw data. They didn't give us the raw data. They just gave us the scores. For example, let's say you take the SAT, which is a test. Okay. The SAT I look up for the manual, SAT manual, and I can see that the reliability is 0.90.

I can look at the validity and find out what the validity is, and from the 0.90 reliability they'll give me what I call -- what they call a standard error measurement, which may be two-and-a-half points.

It's kind of like when you look at a -- they predict election results, and they say this is accurate within plus-or-minus two percent or something like that. I could show you on the normal curve exactly what that means and how much -- you know, how much error you have.

But when you only have the score of 0.66 the only thing you can say is 34 percent of that score is error. You can estimate obviously the error variance is fairly high. On IQ tests for example the standard error of measurement is about five points and the reliability is usually about 0.90.

So what we need to know is when a person gets a score

for --- on the nystagmus test is that plus-or-minus one-and-a-half points or plus-or-minus two points in terms of errors? Or if you're talking about the one-leg-stand I think there's nine queues or something, whether or not it's plus-or-minus one, or two, or three.

Q Sir, based on the '77 and the '81 studies, is it at this point we just don't know what the potential rate of error is for the performance of those three field sobriety tests? Is that what your opinion is?

A Well, I'm not saying that. I'm saying I can't further define it down to giving you clearly in your mind -- for example, if you're on an SAT test, if you took the SATs, you know, the mean is 1,000. Okay? I can tell you that if you take that test and -- because I know the standard error of measurement.

If you took that test again your chances of making between -- if you made 1,000, between 950 and 1,050, you will do that on the average two-thirds of the time. Between 900 and 1,100 I can tell you that you're going to make between 900 and 1,100 the next time you take the test.

If I want to go further, I can go to 98 percent confidence rate. So I can -- that's a better way of giving the data so I can tell you the chances for example if this person scored a certain thing on the walk-and-turn I can tell you how much confidence that you have in that.

With not giving us the raw data I can only tell you that there is 36 percent error variance, and that doesn't mean that it's 36 percent of the time wrong. It only means 36 percent of that score is error variance and the 74 -- 64 I mean is true variance.

Q Okay. And the same for the HGN test and the walk-and --

A Yes. That's correct. That's correct.

Q The one-leg-stand test we can do the same thing? We can have this value of the test?

A Yes. If they give us the raw data then we would. We've asked for it, but they don't.

Q And did that -- came out in your testimony earlier when we were talking about that. You were showing us how on the charts that you went through with us earlier showing on the test/retest reliabilities.

You commented on how -- like for instance on page 35 of the '81 study we circled how the nystagmus score for the officer -- when the same officer did the test two weeks apart he had a reliability of 0.66, which indicated that there was a 34 percent --

A Error variance.

Q Error variance. Okay. So that has been determined then in all these cases.

A Yes. That's right.

Q Okay. I am tracking now. Thank you.

A That's right. It's just -- what I'm saying is that's just -- you know, it's just hard in your -- to understand what does that mean. You know, I know it's high, but just how high is that, you know.

Q I understand. Thank you.

MR. MARONE: Judge, I have no further questions right now.

THE COURT: Ms. Natapoff?

MS. NATAPOFF: I have nothing else, Your Honor.

THE COURT: All right. Doctor, thank you very much. Step down. If we have any more questions to ask of you before we recess and you are still here we will pull you back up and ask you.

THE WITNESS: All right. I think I'm flying out tomorrow. So I'll probably be here.

THE COURT: Okay, sir.

(Witness excused.)

THE COURT: Ms. Natapoff?

MS. NATAPOFF: Your Honor, would it be possible to take a 10-minute recess before we put Mr. Brull on?

THE COURT: Not only is it possible, but it probably will be insufferable not to allow you to do so. We will take 10-minute recess.

(Whereupon, a brief recess was taken.)

Whereupon,