Field Sobriety Tests: How Basic Science Proves They Have Little Power to Tell Impaired From Sober

By Greg Kane, M.D.

Over and over, field sobriety test validation data proves FSTs have no meaningful power to identify impairment.

In the last article we discovered that the so called "accuracy" field sobriety test proponents claim validates FSTs is actually a technical mathematical statistic with weird, non-intuitive properties.[1]

Using exactly the formula NHTSA validation contractors use, we saw that a phony heads-I-win-tails-you-lose coin toss has an "accuracy" of 0%. **And** 1%. **And** 50%. **And** 82%. The NHTSA's so called "accuracy" is actually a triple-percentage created by a formula that passes the roadside sobriety tests' two fundamental accuracy percentages through a third percentage, the percentage of impaired drivers in the validation study group. The so called "accuracy" NHTSA contractors discover is highly dependent on the percentage of impaired drivers the contractors choose to study.

What's more, using real world roadside sobriety test data, we saw that the NHTSA puts forward a so called "accuracy" statistic that expects us to believe a coin toss spots impaired drivers with 82% accuracy. A **coin toss** is 82% accurate? That's crazy. There must be something wrong with the NHTSA's analysis.

There is. Science knows the NHTSA's so called "accuracy" statistic doesn't mean what you think it means. In this article I'll explain science's standard fix for this problem.

Once we know the fix, we'll apply standard scientific techniques to data from field sobriety test validation studies, and discover exactly how well FSTs identify impaired drivers. The answer is that FSTs have no meaningful power to identify impairment.

The problem

The probability that a roadside sobriety test's answer is correct - the percentage chance that a driver who fails the test is actually impaired - depends on the percentage chance, before the test, that the driver was impaired. This confusing behavior is created by NHTSA validation contractors' choice of a so called "accuracy" statistic that applies percentages to percentages.

Let's see if I can make the NHTSA's triple-percentage "accuracy" intuitive for you.

You're on a game show. The host shows you two tubs. The tubs are full of balls. One tub has white balls. The other tub has red balls.

Half the white balls have thousand-dollar checks hidden inside. Ninety-eight percent of the red balls have thousand dollar checks inside.

If you got to reach into a tub and pick a ball, which tub would you choose? The red tub, obviously. From that tub you'd get a check 98% of the time. From the white tub you'd get a check only 50% of the time.

Sorry, you don't get to choose the tub. Instead, the host pours red balls from the red tub and white balls from the white tub into a big sack. You get to pick from the sack. You close your eyes, reach in, and pull out a ball.

What's your chance of winning money?

The answer is, "It depends." If **all** the balls in the sack are red balls, your chance is 98%. If all the balls are white, your chance is 50%. If the balls are a mixture of white and red, then your chance is somewhere between 50% and 98%. The more red balls in the sack, the closer your chance is to the red percentage, 98%. The more white balls, the closer your chance is to the white percentage, 50%.

Your chance of picking a money ball depends on the two fundamental per-

centages - the white ball and red ball "has-money" percentages, AND on a third percentage, the percentage of balls in the sack that are red (or white).

That's more or less how roadside sobriety tests work. Sobriety tests have two fundamental percentages, the Impaired Driver Accuracy and the Innocent Driver Accuracy. No matter what group of drivers you chose to test, when the test is done on impaired drivers it will give the correct answer 96% of the time. No matter what group you chose, on innocent drivers, the test's answer will be correct 56% of the time.[2]

But at the roadside, officers don't test only impaired drivers or only innocent drivers. They test some mixture of drivers, some impaired, some innocent. Just as pulling out a ball with a check depends on the percentage of red and white money balls in the mix, the percentage of time the roadside sobriety test gives the correct answer depends on the percentage of impaired (or innocent) drivers in the mix of drivers the officer has to do the roadside sobriety test on.

That fact has real world implications. The percentage of impaired drivers at a Monday morning sobriety checkpoint is different from the percentage of impaired drivers stumbling out of a bar Saturday night. Because of that difference, the accuracy of an FST/ officer's arrest decision - the percentage of the time that decision is correct - is different when the officer has arrested a driver at a stop-all-drivers sobriety check point, than it is when he's arrested a driver leaving a bar. How different? Way different.

The fix

Science's solution - I can't decide if this is clever or pedestrian - is just to apply the roadside test's two fundamental accuracy percentages to every possible percentage of impaired drivers in any study group. (The well known formula for this is given in the *PPV Sidebar*.) Then, when you need to find the accuracy of the roadside sobriety test for any one particular driver, you just figure out that driver's pre-test probability of impairment and plug that number into the positive predictive value formula. Out the other end comes the percent chance that one particular driver's arrest was correct, the percent chance that one driver, who failed the roadside sobriety test, was actually impaired.

The result

Why should you care? Two reasons.

First, the scientific positive predictive value formula has the powerful ability to show exactly how much "guilt" is contained in a roadside sobriety test's "Yes" answer. You can do it yourself. Just compare the pre-test probability of impairment with the answer that comes out of the formula, the "post-test probability." The difference between those two numbers is how much certainty about "guilt" the test added.

Look at the *Watch How the Percentages Change Sidebar*, at the results for the Colorado Validation Study. Drivers with a 1% chance of impairment, when they failed the Colorado Validation Study roadside sobriety test, still had only a 4% chance of being impaired. A driver with a 5% likelihood of impairment, after an "arrest" answer, had only a 16% chance of being impaired. All across the table, the FST added only a few percentage points to the probability of guilt - never enough to turn unlikely into likely.

That's the point. It's not that you have to do a fancy positive predictive value mathematical accuracy statistic calculation for every DUI defendant. The point is that roadside sobriety tests

Continued on page 61

The PPV formula

Scientific tests are not perfect. You do the test. You get an answer. The probability that answer is correct is called the test's "positive predictive value." The formula is:

$$PPV = \frac{\text{sensitivity} \times \text{prevalence}}{(\text{sensitivity} \times \text{prevalence}) + (1-\text{specificity})(1-\text{prevalence})}$$

As you see, to calculate the PPV of a test you need three numbers:

The sensitivity of the test (the Impaired Driver Accuracy)

The specificity of the test (the Innocent Driver Accuracy)

The prevalence of impairment in the group tested

The sensitivity and specificity are invariant, fundamental properties of the test. This is not theoretical stuff. You can use published field sobriety test validation data to calculate the sensitivity and specificity of officer arrest decisions. When you do that, the sensitivity and specificity numbers you calculate apply to the group of drivers in the study - and to tea totaling Sunday school teachers, and drunken bikers and everyone else you can find to do a field sobriety test on.* Very powerful. Very cool.

The prevalence is a property of the test subject. For details of how science deals with prevalence, see the main article.

To learn more, and for thousands of references in the scientific literature, Google "'positive predictive value' prevalence," or "sensitivity specificity predictive value," etc.

* This sentence describing how statistics work presumes the physical test works the same on different people. Whether the FST balance, and walk and turn test is in fact as accurate on innocent 70 year olds as it is on innocent 20 year olds is a question NHTSA validation contractors fail to answer.

Percentages

Watch how the percentages change. Let's use the positive predictive value formula[i] to calculate the power of various tests to identify guilt. We'll start with the standard formulas for sensitivity (Impaired Driver Accuracy) and specificity (Innocent Driver Accuracy). Once we have those numbers, we'll calculate the positive predictive value, the PPV - the accuracy of the officer's arrest decision - for a pre-test probability of impairment of 1%. Then we'll calculate the PPV again, for a pre-test probability of 5%. Then 10%, 20%, all the way up to 100%.

Keep an eye on how much the percentages change. A test with lots of impairment-finding power will change the percentages a lot. If a test doesn't change the percentages much, it has little power to find impairment.

We'll start with a coin-toss test, because we know coin tosses have no power to find impairment. The PPV will say that - if the PPV works.

Then we'll look at the PPV of a highly accurate alcohol blood test. The PPV will confirm that this accurate test's Yes means Yes, and its No means No - if the PPV works.

Finally, using real-world validation study data, we'll discover that roadside sobriety tests have shockingly little power to identify impairment.

Keep an eye on how much the percentages change. You'll learn a lot.

#1 Time Efficient-FST: Let's start with the phony Time Efficient Field Sobriety Test from the last article,[ii] the coin toss that says everyone is guilty.

		Roadside Test		
		sober	impaired	
Driver BAC	impaired	0	5	**100%** = Sensitivity = the accuracy of the test on impaired drivers
	sober	0	5	**0%** = Specificity = the accuracy of the test on sober drivers

Now that we know the sensitivity and specificity of the test, the PPV formula tells us:

When the pretest probability of impairment is this:	1	**5**	10	20	30	**40**	50	60	70	80	**90**	95	100
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
A driver who fails a **Time Efficient FST** has this % chance of being impaired:	1	**5**	10	20	30	**40**	50	60	70	80	**90**	95	100

Notice the percentages don't change. If a driver has a 5% pre-test chance of being impaired, a positive Time Efficient Field Sobriety Test changes that percent chance to . . . 5%! 40% to 40%. 90% to 90%. No change at all.

Now you see why science uses PPV analysis. The NHTSA validation contractors' so called "accuracy" statistic told us this phony coin-toss was "83% accurate."[iii] A *coin toss* is 83% accurate? That's crazy. The NHTSA's analysis must be wrong. It is. Science's fix is the PPV.

Science's standard PPV analysis gives the result we know is right - a test with no power to tell impaired from innocent should not change our assessment of the probability of impairment. And PPV analysis shows this test doesn't. That's why scientists use PPV analysis. It works.

#2 Alcohol Blood Test: Now let's analyze an imaginary **alcohol blood test**, a highly accurate chemical laboratory test:

		Alcohol Blood Test		
		sober	impaired	
Driver BAC	impaired	0	999	**99%** = Sensitivity = the accuracy of the test on impaired drivers
	sober	9999	1	**99.99%** = Specificity = the accuracy of the test on sober drivers

Now that we know the sensitivity and specificity of the test, the PPV formula tells us that:

When the pretest probability of impairment is this:	**1**	**5**	10	20	30	40	50	60	70	80	90	95	100
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
Any driver who fails a this **blood alcohol test** has this % chance of being impaired:	**99**	**100**	100	100	100	100	100	100	100	100	100	100	100

[i] See, *PPV* sidebar.

[ii] Greg Kane, M.D., *Field Sobriety Tests: Percentages of Percentages, Why "Validation" Studies Fail to Validate*, TRIAL TALK, Aug./Sept. 2006 at 31.

[iii] Greg Kane, M.D., *Field Sobriety Tests,* TRIAL TALK, Aug./Sept. 2006

Again we get exactly the result we expect. 5% becomes 100%. Even drivers with very low pre-test probabilities of impairment are correctly and reliably identified by this highly accurate test. Yes means Yes. No means No.

This brings up a key point. For highly accurate tests, you don't need to worry about the pre-test probability of impairment. This imaginary blood alcohol test is so powerful that, when drivers fail it, even drivers with very low pre-test probabilities are overwhelmingly likely to be impaired. We'll see in a minute that FSTs and officer arrest decisions work very differently.

By the way, I set up this example to help you understand field sobriety test accuracy statistics. What the exact sensitivity and specificity of real-world blood tests are, I don't know; and for the purposes of explaining to you how FSTs work, I don't care.

#3 Colorado Validation Study, BAC > 0.05%

Now lets use the data NHTSA validation contractors imagine proves officer arrest decisions are 93% accurate. (Validation studies generally measure not FSTs but officer arrest decisions.) Here are the actual results from the NHTSA-funded **Colorado Validation Study**, testing the so called "accuracy" with which an officer arrest identifies driver BACs as greater than 0.05%.[iv]

		Roadside Test		
		sober	impaired	
Driver BAC	impaired	21	163	**89%** = Sensitivity = the accuracy of the test on impaired drivers
	sober	38	12	**76%** = Specificity = the accuracy of the test on sober drivers

When the pretest probability of impairment is this:	1	**5**	10	20	30	40	50	60	**70**	80	90	95	100
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
A driver who fails an **officer arrest/release decision** has this % chance of being impaired:	4	**16**	29	48	61	71	79	85	90	94	97	99	100

Bad news

How much implication of guilt does this roadside test add? Not much. Look at how little the percentages change. Drivers with a 1% chance of impairment, when they fail this Colorado Validation Study FST, still have only a 4% chance of being impaired. A driver with a 5% likelihood of impairment, after an "arrest" answer, has only a 16% chance of being impaired.

And at higher levels of certainty, 70% pre-test becomes only 90%. The test adds only 20 percentage points. Looked at the other way, if the jury is 90% certain the driver was impaired, only 22% of their certainty [=20/90] came from the roadside test, the rest came from a presumption of guilt.

That's the point. It's not that you have to do a PPV calculation for every DUI defendant. It's that standard scientific analysis of NHTSA validation study data proves roadside tests add very little probability of guilt to anyone. FSTs are very weak tests. Scientifically, FSTs do not work.

Many validation studies, always the same result - FSTs do not work.

This FSTs don't work answer is not an anomaly found just in the Colorado Validation Study. Over and over, validation studies have proven the same result.

1977, FST[v]

		Roadside Test		
		sober	impaired	
Driver BAC	impaired	21	163	**89%** = Sensitivity = the accuracy of the test on impaired drivers
	sober	38	12	**76%** = Specificity = the accuracy of the test on sober drivers

When the pretest probability of impairment is this:	1	**5**	10	20	30	40	50	60	**70**	80	90	95	100
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
A driver who fails an **officer arrest/release decision** has this % chance of being impaired:	4	**16**	29	48	61	71	79	85	90	94	97	99	100

[iv] Marcelline Burns and Ellen W. Anderson, *A Colorado Validation Study of the Standardized Field Sobriety Test (SFST) Battery*, Colo. Dep't of Transp., 1995.

[v] M. Burns and H. Moskowitz. *Psychophysical Tests for DWI Arrests*, NHTSA, DOT-HS-5-01242 (1977).

Florida Validation Study, BAC > 0.08%[vi]

		Roadside Test		
		sober	impaired	
Driver BAC	impaired	9	197	**96%** = Sensitivity = the accuracy of the test on impaired drivers
	sober	41	9	**82%** = Specificity = the accuracy of the test on sober drivers

When the pretest probability of impairment is this:	1	**5**	10	20	30	40	50	60	**70**	80	90	95	100
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
A driver who fails an **officer arrest/release decision** has this % chance of being impaired:	**5**	22	37	57	69	78	84	89	93	96	98	99	100

1998, Officer's Estimated BAC 0.10% 1998[vii]

		Roadside Test		
		sober	impaired	
Driver BAC	impaired	4	210	**98%** = Sensitivity = the accuracy of the test on impaired drivers
	sober	59	24	**71%** = Specificity = the accuracy of the test on sober drivers

When the pretest probability of impairment is this:	1	**5**	10	20	30	40	50	60	**70**	80	90	95	100
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
A driver who fails an **officer arrest/release decision** has this % chance of being impaired:	3	15	27	46	59	69	77	84	89	93	97	98	100

[vi] Burns, *A Florida Validation Study of the Standardized Field Sobriety Test (S.F.S.T.) Battery*, NHTSA (1997).

[vii] Jack Stuster and Marcelline Burns, *Validation of the Standardized Field Sobriety Test Battery at BACs Below 0.10*., NHTSA, 1998 page 18, figure 4.

Continued from Page 58

have been studied and restudied and studied again. And over and over and over standard scientific analysis of the results prove exactly the same thing. Roadside sobriety tests have no meaningful power to turn unlikely into likely - no meaningful power to identify impairment. FSTs do not work.

The second reason you should care about PPV science is that, as you've noticed, the probability of guilt revealed by the test is highly variable. Think of it this way. In a football game, the guy with the ball just ran twelve yards. Where's the ball now? I don't know. You don't know. It could be on the twelve yard line. It could be way down in the other end zone. Where the ball is after a twelve yard run depends on where the ball started.

Roadside sobriety tests work the same way. The test was positive - that moves the probability a few percentage points forward. Where's the probability now? I don't know. You don't know. The defendant might be 95% likely to be impaired. The defendant might be 4% likely to be impaired. When a test moves the probability forward only a few yards, err, percentage points, the place the probability ends up depends mostly on where the probability was before the test.

This annoying property does not happen because I'm being pedantic about scientific protocol. The science is what it is. The meaninglessness of the roadside sobriety test in the absence of the pretest probability of impairment happens because of the nature of the roadside sobriety test itself. Roadside sobriety tests are very weak tests. They have little power to tell impaired from sober. Weak tests don't move the probability far. Where it ends up depends mostly on where it started.

To know where the football is, now that you know there was just a twelve-yard run, you must know where the ball was before the run. Any interpretation of where the ball is now that fails to include an assessment of where it was before the run is really nothing but a guess.

To know how likely a driver is to have been impaired, now that you know he failed an FST, you must know where he started. You must know how likely he was to have been impaired before the test. Any interpretation of a field sobriety test or officer arrest decision that fails to include a *scientific* assessment of the driver's pre-test probability of impairment is nothing but a guess, our modern pseudo-science equivalent of pushing witches in the village pond.

Finally, you've maybe noticed that even though we're talking about reports styled as **field sobriety test** validation studies I'm careful to say "roadside sobriety test" accuracy, and not "field sobriety test" accuracy. There's a

reason. The truth is "field sobriety test" validation studies don't gather data on or do statistical analysis of field sobriety tests. They actually measure something else - the accuracy of officer arrest decisions. They do not gather data on or do a statistical analysis of what contribution, if any, FSTs make to officers' arrest "accuracy."

That means that the increments of guilt revealed by the PPV science are not increments created by the FST. They are increments created by everything the officer does after he's decided to include the driver in the validation study. And that means the increments are nothing but the upper limits of what FSTs *might* contribute - if they contribute anything at all. I'll talk more about this in a later article.

If FSTs don't work, how come do NHTSA validation studies validate them as "extremely accurate"? In the next article in this series, I'll show you two simple statistical tricks that let NHTSA validation contractors "validate" **any** FST.

Greg Kane, MD owns Med-mal Experts, Inc., an Englewood based consulting firm that reviews medical malpractice claims and refers attorneys across America to physician expert witnesses in all specialties. Reach him at 303-741-0993 or FST@medmalExperts.com or visit www.medmalExperts.com.

End Notes

[1] Greg Kane, M.D., *Field Sobriety Tests: Percentages of Percentages, Why "Validation" Studies Fail to Validate*, TRIAL TALK, Aug./Sept. 2006 at 31.

[2] If, that is, the roadside sobriety test is actually a dichotomous scientific test that gives repeatable results; data from Colorado Validation Study, Appendix IV, results for BAC 0.10%. 2006 at 31.