UNIVERSITY OF CENTRAL ARKANSAS™ | COLLEGE OF BUSINESS

Assessing the Potential Disparate Impacts of Hair Drug Testing Among Commercial Motor Vehicle Drivers

February 7, 2020

Prepared for:

The Alliance for Driver Safety and Security (Trucking Alliance)
Washington, D.C.

Prepared by:

Joe Cangelosi, Ph.D.
Professor of Marketing
University of Central Arkansas

Doug Voss, Ph.D.
Professor of Logistics and Supply Chain Management
University of Central Arkansas

Assessing the Potential Disparate Impacts of Hair Drug Testing Among Commercial Motor Vehicle Drivers

INTRODUCTION

The Alliance for Driver Safety and Security (The Trucking Alliance) asked Drs. Doug Voss and Joe Cangelosi (Researchers) to assess whether hair drug tests disparately impact minority groups when compared to urine drug tests.

Researchers utilized two methods to assess disparate impact. First, the **"Four-Fifths Rule"** is defined in the Code of Federal Regulations, Title 29 - Uniform Guidelines for Employee Selection Procedures, as "a selection rate for any race, sex or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact."

In other words, disparate impact is assumed if any ethnic group does not pass at a rate of at least 80% of the rate of the ethnic group with the highest passing rate.

Second, Researchers utilized **"chi-square difference tests"** to assess whether significant differences exist between ethnic groups within each test (e.g. whether a significant difference exists between ethnic groups for urine tests and, separately, whether a significant difference exists between ethnic groups for hair tests).

Chi-square results would indicate disparate impact if no significant between-group differences exist for urine testing but do exist for hair testing. This would imply that the groups' urine test pass/fail rate is statistically equivalent, but the groups' hair test pass/fail rate is significantly different. Alternatively, chi-square results would indicate equal treatment if significant between-group differences exist for both/neither urine and hair tests. This would imply that the groups pass/fail rates are statistically equivalent/different irrespective of testing procedure.

Researchers were independently provided with paired urine and hair pre-employment drug screen results from three (3) commercial trucking companies for the years 2017-2019. Two (2) companies provided results from 2017, three (3) provided results from 2018, and one (1) provided results from 2019. To provide anonymity to the carriers and their drivers, our analysis focused on two (2) different time periods: 2017-2019 combined and 2018 in isolation. Sample sizes for each test are as follows:

- 2017-2019 urine test: n = 73,176
- 2017-2019 hair test: n = 72,023
- 2018 urine test: n = 39,517
- 2018 hair test: n = 38,900

Given a margin of error = 1%, and a confidence level = 99%, a sample size of 16,641 is required to generalize results across the broader U.S. truck driver population. Sample sizes presented above exceed this threshold and results can be generalized Nationally.[1]

RESULTS[2]

FOUR-FIFTHS RULE

TABLE 1
2017-2019 URINE TEST RESULTS

ETHNIC GROUP	PASSED	FAILED	TOTAL	PERCENT PASSED	PERCENT OF HIGHEST PASSING RATE (ASIAN)
AM. INDIAN	753	6	759	99.2%	99.6%
ASIAN	1802	7	1809	99.6%	100.0%
BLACK	28632	294	28926	99.0%	99.4%
HAWAII/PACIFIC ISLANDER	276	2	278	99.3%	99.7%
HISPANIC	8191	44	8235	99.5%	99.9%
MULTIPLE	1777	25	1802	98.6%	99.0%
NOT SPECIFIED	8327	144	8471	98.3%	98.7%
WHITE	22664	232	22896	99.0%	99.4%
TOTAL	72422	754	73176	99.0%	99.4%

*Pearson chi-square = 67.52; p = 0.00; n = 73,176

Table 1 details 2017-2019 urine test results. Ninety nine percent (99%) of drivers in the Asian ethnic group passed their pre-employment drug screens. To comply with the Four-Fifths Rule, every other ethnic group must pass at a rate equal to 80% of this figure (99% x 80% = 79%). Drivers who chose not to report their ethnic group ("not specified") passed at the lowest rate, which was 98.7% of the ethnic group with the highest passing rate. This exceeds the required Four-Fifths Rule 79% threshold.

[1] Cangelosi and Voss (2019), "An Examination of the Geographical Correlation Between Commercial Motor Vehicle Drivers," prepared for The Alliance for Driver Safety and Security.

[2] Appendix A details chi-square cross tabulation results. Appendix B provides failure rates for each ethnic group and rank orders failure rates for each group in each sample. Appendix C provides a chi-square test tutorial.

TABLE 2
2017-2019 HAIR TEST RESULTS

ETHNIC GROUP	PASSED	FAILED	TOTAL	PERCENT PASSED	PERCENT OF HIGHEST PASSING RATE (ASIAN)
AM. INDIAN	709	48	757	93.7%	97.0%
ASIAN	1739	61	1800	96.6%	100.0%
BLACK	26329	2215	28544	92.2%	95.5%
HAWAII/PACIFIC ISLANDER	258	17	275	93.8%	97.1%
HISPANIC	7699	452	8151	94.5%	97.8%
MULTIPLE	1655	139	1794	92.3%	95.5%
NOT SPECIFIED	7149	925	8074	88.5%	91.7%
WHITE	21678	950	22628	95.8%	99.2%
TOTAL	67216	4807	72023	93.3%	96.6%

*Pearson chi-square = 624.6; p = 0.000; n = 72,023

Table 2 details 2017-2019 hair test results. Ninety six percent (96%) of drivers in the Asian ethnic group passed their pre-employment drug screens. To comply with the Four-Fifths Rule, every other ethnic group must pass at a rate equal to 80% of this figure (96% x 80% = 77%).

Drivers represented by the Black and Multiple ethnic groups passed at the lowest rate, which was 95.5% of the ethnic group with the highest passing rate. This exceeds the required 77% Four-Fifths Rule threshold.

TABLE 3
2018 URINE TEST RESULTS

ETHNIC GROUP	PASSED	FAILED	TOTAL	PERCENT PASSED	PERCENT OF HIGHEST PASSING RATE (HAW/PAC IS.)
AM. INDIAN	367	2	369	99.5%	99.5%
ASIAN	847	5	852	99.4%	99.4%
BLACK	15345	166	15511	98.9%	98.9%
HAWAII/PACIFIC ISLANDER	144	0	144	100.0%	100.0%
HISPANIC	4022	23	4045	99.4%	99.4%
MULTIPLE	851	17	868	98.0%	98.0%
NOT SPECIFIED	6043	118	6161	98.1%	98.1%
WHITE	11455	112	11567	99.0%	99.0%
TOTAL	39074	443	39517	98.9%	98.9%

*Pearson chi-square = 59.43; p = 0.000; n = 39,517

Table 3 details 2018 urine test results. One hundred percent (100%) of drivers in the Hawaii/Pacific Islander ethnic group passed their pre-employment drug screens. To comply with the Four-Fifths Rule, every other racial group must pass at a rate equal to 80% of this figure (100% x 80% = 80%).

Drivers who chose not to report their ethnic group ("not specified") passed at the lowest rate, which was 98.1% of the ethnic group with the highest passing rate. This exceeds the required 80% Four-Fifths Rule threshold.

TABLE 4
2018 HAIR TEST RESULTS

ETHNIC GROUP	PASSED	FAILED	TOTAL	PERCENT PASSED	PERCENT OF HIGHEST PASSING RATE (HAW/PAC IS.)
AM. INDIAN	344	25	369	93.2%	96.6%
ASIAN	812	35	847	95.9%	99.3%
BLACK	14163	1173	15336	92.4%	95.7%
HAWAII/PACIFIC ISLANDER	136	5	141	96.5%	100.0%
HISPANIC	3780	239	4019	94.1%	97.5%
MULTIPLE	795	71	866	91.8%	95.1%
NOT SPECIFIED	5110	731	5841	87.5%	90.7%
WHITE	11000	481	11481	95.8%	99.3%
TOTAL	36140	2760	38900	92.9%	96.3%

*Pearson chi-square = 438.14; p = 0.000; n = 38,900

Table 4 details 2018 hair test results. Ninety six percent (96%) of drivers in the Hawaii/Pacific Islander ethnic group passed their pre-employment drug screens. To comply with the Four-Fifths Rule, every other ethnic group must pass at a rate equal to 80% of this figure (96% x 80% = 77%). Drivers who chose not to report their ethnic group ("not specified") passed at the lowest rate, which was 90.7% of the ethnic group with the highest passing rate. This exceeds the required 77% Four-Fifths Rule threshold.

CHI-SQUARE DIFFERENCE TESTS

Chi-square results are presented as footnotes below tables 1-4. Significant differences across ethnic groups were found for urine tests across all years and 2018 in isolation. Significant differences across ethnic groups were found for hair tests across all years and 2018 in isolation. Chi-square results indicate equal treatment if significant between-group differences exist for both urine and hair testing. This indicates the groups pass/fail rates are statistically different for urine testing and are also statistically different for hair testing. Irrespective of testing procedure, ethnic groups' drug test results are significantly different.

CONCLUSIONS

Utilizing independently provided urine and hair pre-employment drug screen data, Researchers were unable to find disparate impacts of hair testing among the ethnic groups analyzed. Results for each test in each sample met the required Four-Fifths Rule threshold.

Chi-square tests independently examine urine and hair tests. Chi-square results indicate that the proportion of drug test failures (positives) are higher for hair testing across all ethnic groups but pass/fail rates are significantly different irrespective of testing method.

Given these findings, Researchers find no disparate impact among ethnic groups by testing method.

APPENDIX A
CHI-SQUARE CROSS-TABULATION RESULTS[3]

URINE TEST
2017 - 2019

		Asian	Black	Did Not Disclose	Hispanic	Islander	Am. Indian	White	Multiple	TOTALS
TOTALS	Negative-OBS	1802	28632	8327	8191	276	753	22664	1777	72422
	Negative-EXP	1790	28628	8384	8150	275	751	22660	1783	
	diff-SQ	0.08	0.00	0.38	0.20	0.00	0.00	0.00	0.02	
	Positive-OBS	7	294	144	44	2	6	232	25	754
	Positive-EXP	19	298	87	85	3	8	236	19	
	diff-SQ	7.27	0.06	36.85	19.67	0.26	0.42	0.07	2.23	
	Totals	1809	28926	8471	8235	278	759	22896	1802	73176
	Comp Chi-SQ	67.52								
	Deg Freedom	7								
	Crit Chi-SQ	**14.07**	at 95% confidence							

3 OBS: Actual Observed results of the pre-employment testing.
EXP: Expected results of the pre-employment test if there were no statistically different failure rates across ethnic groups.
Computed Chi-SQ: the actual computed value of Chi-Square given the differences among failure rates across ethnic groups.
Critical Chi-SQ: the minimum value of Chi-Square to indicate statistically significances across ethnic groups at a 95% confidence level.

If the computed Chi-Square is larger than the Critical Chi-Square, then we conclude that statistically significant differences exist across groups.

HAIR TEST
2017 - 2019

		Asian	Black	Did Not Disclose	Hispanic	Islander	Am. Indian	White	Multiple	TOTALS
TOTALS	Negative-OBS	1739	26329	7149	7699	258	709	21678	1655	67216
	Negative-EXP	1680	26639	7535	7607	257	706	21118	1674	
	diff-SQ	2.08	3.61	19.79	1.11	0.01	0.01	14.86	0.22	
	Positive-OBS	61	2215	925	452	17	48	950	139	4807
	Positive-EXP	120	1905	539	544	18	51	1510	120	
	diff-SQ	29.11	50.41	276.66	15.56	0.10	0.13	207.83	3.10	
	Totals	1800	28544	8074	8151	275	757	22628	1794	72023
	Comp Chi-SQ	624.60								
	Deg Freedom	7								
	Crit Chi-SQ	14.07	at 95% confidence							

URINE TEST
2018

		Asian	Black	Did Not Disclose	Hispanic	Islander	Am. Indian	White	Multiple	TOTALS
TOTALS	Negative-OBS	847	15345	6043	4022	144	367	11455	851	39074
	Negative-EXP	842	15337	6092	4000	142	365	11437	858	
	diff-SQ	0.02	0.00	0.39	0.12	0.02	0.01	0.03	0.06	
	Positive-OBS	5	166	118	23	0	2	112	17	443
	Positive-EXP	10	174	69	45	2	4	130	10	
	diff-SQ	2.17	0.36	34.67	11.01	1.61	1.10	2.41	5.43	
	Totals	852	15511	6161	4045	144	369	11567	868	39517
	Computed Chi-SQ	59.43								
	Deg Freedom	7								
	Critical Chi-SQ	14.07	at 95% confidence							

HAIR TEST
2018

		Asian	Black	Did Not Disclose	Hispanic	Islander	Am. Indian	White	Multiple	TOTALS
TOTALS	Negative-OBS	812	14163	5110	3780	136	344	11000	795	36140
	Negative-EXP	787	14248	5427	3734	131	343	10666	805	
	diff-SQ	0.80	0.51	18.47	0.57	0.19	0.00	10.43	0.11	
	Positive-OBS	35	1173	731	239	5	25	481	71	2760
	Positive-EXP	60	1088	414	285	10	26	815	61	
	diff-SQ	10.48	6.62	241.83	7.47	2.50	0.05	136.61	1.49	
	Totals	847	15336	5841	4019	141	369	11481	866	38900
	Computed Chi-SQ	438.14								
	Deg Freedom	7								
	Critical Chi-SQ	14.07	at 95% confidence							

APPENDIX B
COMPARING URINE AND HAIR TEST FAILURE RATES

ETHNIC GROUP	2018 Urine	2018 Hair	Difference	2017-2019 Urine	2017-2019 Hair	Difference
AM. INDIAN	0.5%	6.8%	**6.2%**	0.8%	6.3%	**5.6%**
ASIAN	0.6%	4.1%	**3.5%**	0.4%	3.4%	**3.0%**
BLACK	1.1%	7.6%	**6.6%**	1.0%	7.8%	**6.7%**
HAWAII/PACIFIC ISLANDER	0.0%	3.5%	**3.5%**	0.7%	6.2%	**5.5%**
HISPANIC	0.6%	5.9%	**5.4%**	0.5%	5.5%	**5.0%**
MULTIPLE	2.0%	8.2%	**6.2%**	1.4%	7.7%	**6.4%**
NOT SPECIFIED	1.9%	12.5%	**10.6%**	1.7%	11.5%	**9.8%**
WHITE	1.0%	4.2%	**3.2%**	1.0%	4.2%	**3.2%**
AVERAGE	1.0%	6.6%	**5.7%**	0.9%	6.6%	**5.6%**

Appendix B compares urine and hair test failure rates across ethnic groups for both samples.

The table above illustrates that each ethnic group had a higher hair test failure rate. Tables on the next page rank order the failure rate differences for the two samples then draw comparisons.

ETHNIC GROUP	2018 Urine	2018 Hair	Difference
NOT SPECIFIED	1.9%	12.5%	**10.6%**
BLACK	1.1%	7.6%	**6.6%**
MULTIPLE	2.0%	8.2%	**6.2%**
AM. INDIAN	0.5%	6.8%	**6.2%**
AVERAGE	1.0%	6.6%	5.7%
HISPANIC	0.6%	5.9%	**5.4%**
HAWAII/PACIFIC ISLANDER	0.0%	3.5%	**3.5%**
ASIAN	0.6%	4.1%	**3.5%**
WHITE	1.0%	4.2%	**3.2%**

ETHNIC GROUP	2017-2019 Urine	2017-2019 Hair	Difference
NOT SPECIFIED	1.7%	11.5%	**9.8%**
BLACK	1.0%	7.8%	**6.7%**
MULTIPLE	1.4%	7.7%	**6.4%**
AM. INDIAN	0.8%	6.3%	**5.6%**
AVERAGE	0.9%	6.6%	5.6%
HAWAII/PACIFIC ISLANDER	0.7%	6.2%	**5.5%**
HISPANIC	0.5%	5.5%	**5.0%**
WHITE	1.0%	4.2%	**3.2%**
ASIAN	0.4%	3.4%	**3.0%**

The above tables rank order the ethnic groups in each sample by the difference between hair and urine test failure rates. The rank order across samples was very similar.

The Not Specified group had the greatest disparity in both the 2018 (10.6%) and 2017-2019 (9.8%) samples. Blacks had the second greatest disparity (2018 = 6.6%; 2017-2019 = 6.7%) but were very closely followed by Multiple (2018 = 6.2%; 2017-2019 = 6.4%) and American Indian 2018 (6.2%). The American Indian failure rate in 2018 was higher than the 2018 sample average. The American Indian failure rate for 2017-2019 (5.6%) was roughly equivalent to the sample average. Hispanic (2018 = 5.4%; 2017-2019 = 5.0%), Hawaii/Pacific Islanders (2018 = 3.5%; 2017-2019 = 5.5%), Asians (2018 = 3.5%; 2017-2019 = 3.0%) and Whites (2018 = 3.2%; 2017-2019 = 3.2%) had lower than average failure rates for both samples.

APPENDIX C
CHI-SQUARE TUTORIAL

- The Chi-Square (CS) test is a test of "Goodness of Fit" across several distributions in an analysis.
- The test allows for the observation of two variables simultaneously.
- The test involves several things:
    - 2 discrete nominal scale variables each containing multiple nominal scale values.
        - FOR EXAMPLE: in our analysis of the possible differences in the drug failure rates for pre-employment commercial drivers, we have two variables (bivariate):
        - 1) ethnicity—8 ethnic backgrounds or really 7 ethnic backgrounds and one category for those applicants that did not disclose their ethnicity.
        - 2) the results of the pre-employment screening drug test, which had two possible results, positive and negative.

The **bivariate CS test** was utilized to test:

- whether significant differences in positive/negative test results for both the Urine and Hair pre-employment drug tests.
- For each test separately, the test would indicate whether or not there were statistically significant differences in the distribution of the positive and negative test results for each of the 8 ethnic categories.
- For example, if the distribution of positive (failed drug test) and negative (passed drug test) test results was **exactly the same** for all 8 ethnic categories, then the computed CS value would be ZERO; hence the null hypothesis of no differences.

- **Computed Chi-SQ value:**
    - The **Computed CS value** is a function of the differences in the expected (EXP) and observed (OBS) frequencies.
    - For example, the null hypothesis assumes that the distributions of positive and negative results are exactly the same for all 8 ethic categories, for example 97% pass and 3% fail.
    - Under the null hypothesis, the computed CS value is zero.

Critical Chi-SQ value: The CS table is easily accessible via Google or any statistics text.

- The CS table adjusts for the size of the matrix being analyzed, for example, for 8 ethnic groups and 2 possible outcomes (positive and negative test results) the number of degrees of freedom would be 7.
- Hence, the critical CS value is the minimum value of a computed CS necessary for statistical significance, for a given level of confidence.
- For example, at 7 degrees of freedom, the critical CS value is 14.07 or 18.48 at the 99% level of confidence.

- Now, let's look at one of the output tables for drug pretesting in the analysis of the Urine data.

TABLE 1-URINE: 2017, 2018, 2019 (ACTUAL DATA)

	Asian	Black	Did Not Disclose	Hispanic	Islander	Native	White	Multiple	TOTALS
Negative-OBS	1802	28632	8327	8191	276	753	22664	1777	72422
Negative-EXP	1790	28628	8384	8150	275	751	22660	1783	
diff-SQ (CS)	0.08	0.00	0.38	0.20	0.00	0.00	0.00	0.02	
Positive-OBS	7	294	144	44	2	6	232	25	754
Positive-EXP	19	298	87	85	3	8	236	19	
diff-SQ (CS)	7.27	0.06	36.85	19.67	0.26	0.42	0.07	2.23	
Totals	1809	28926	8471	8235	278	759	22896	1802	73176
Comp Chi-SQ	**67.52**								
Deg Freedom	7								
Crit Chi-SQ	**14.07**	at 95% confidence							

NOTE:

- OBS = observed frequencies; in other words, the actual data from the test.
- EXP = expected frequencies; what the frequencies would be if there was no difference in the distribution of results across the 8 ethnic groups, hence the values under the null hypothesis.
- Rarely would the OBS and EXP frequencies be identical, but are they different enough so that we can infer at a level of confidence of 95% or greater, that EXP and OBS are statistically significantly different?
- **The CS test looks at the difference in the OBS and EXP frequencies and computes a CS value for each cell in the matrix, indicated by the red number.**
- Notice that in Table 1 above, that for most of the cells,
  - there is very little difference in the OBS and EXP frequencies.
  - However, note that for respondents that "did not disclose" their ethnic background, the rate of positive tests (OBS=144, EXP=87) is almost 66% greater than for what would be expected if there was no difference in the OBS & EXP frequencies.
  - A similar finding is found for the Hispanic ethnic group, where EXP=44 and OBS=85; hence, the positive or failure rate for the Urine test for Hispanics is 93% less than what would be expected, given there was no difference in the OBS & EXP frequencies.
- The two cells in the matrix containing positive testing for "did not disclose" and Hispanic ethnic groups accounted for almost all of the variation in the OBS & EXP frequencies.
  - The result is a computed CS value of 67.52, which is much more than the critical CS value of 14.07 needed for significant statistical differences in OBS and EXP frequencies at a confidence level of 95%.
- **After looking at the statistical output, we can reason from the table results that there is a statistically significant difference in the pass/fail rates of the 8 ethnic groups, and the differences are concentrated in the "Did Not Disclose" and Hispanic ethnic groups.**

In a 2nd example, let's change the data in the table to illustrate a statistically insignificant difference in OBS and EXP frequencies.

- Suppose the EXP and OBS values had been similar as is illustrated below.
- In such a case, the computed CS value would be only 10.65, and we would conclude that there were no statistical differences across the 8 ethnic groups concerning the rate of failure of the urine test.

EXAMPLE TABLE ILLUSTRATING STATISTICAL INSIGNIFICANCE.

	Asian	Black	Did Not Disclose	Hispanic	Islander	Native	White	Multiple	**TOTALS**
Negative-OBS	1802	28632	8327	8191	276	753	22664	1777	72422
Negative-EXP	1791	28637	8330	8186	275	751	22667	1784	
diff-SQ (CS)	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.03	
Positive-OBS	7	294	87	78	2	6	232	25	731
Positive-EXP	18	289	84	83	3	8	229	18	
diff-SQ (CS)	6.79	0.08	0.10	0.26	0.22	0.33	0.04	2.72	
Totals	1809	28926	8414	8269	278	759	22896	1802	73153
Comp Chi-SQ	**10.65**								
Deg Freedom	7								
Crit Chi-SQ	**14.07**	at 95% confidence							

Notice that all of the cells in the matrix have similar values for EXP & OBS frequencies, in order to illustrate a situation in which the OBS & EXP frequencies were NOT significantly different at the 95% level of confidence.