Table 1. Baseline Characteristics (Weighted) of Three U.S. National Data Samples Aged 16 Years and Older: Drivers from the 2013-14 National Roadside Survey of Alcohol and Drug Use (NRS), U.S. population from the 2013-14 National Survey on Drug Use and Health (NSDUH), Homicide Victims from the 2013 National Violent Death Reporting System (NVDRS)

Data Source	NRS	NSDUH	NVDRS
Characteristics	n = 11,314 $n = 43,465Frequency (%) Frequency (%)$		n = 4,110 Frequency (%)
Age (years)			
16-20	1,069 (11.7)	3,817 (8.8)	489 (11.9)
21-34	3,640 (39.8)	10,414 (24.0)	1,813 (44.1)
35-49	2,310 (25.2)	10,751 (24.7)	997 (24.3)
50-64	1,621 (17.7)	10,755 (24.7)	581 (14.1)
≥ 65	515 (5.6)	7,728 (17.8)	230 (5.6)
Missing	2,158	0	0
Sex			
Male	6,382 (58.3)	20,970 (48.2)	3,340 (81.3)
Female	4,566 (41.7)	22,495 (51.8)	770 (18.7)
Missing	365	0	0
Race			
White	4,952 (55.0)	28,458 (65.5)	1,229 (29.9)
Black	2,196 (24.4)	5,131 (11.8)	2,306 (56.1)
Hispanic	1,074 (11.9)	6,622 (15.2)	376 (9.1)
Others	776 (8.6)	3,254 (7.5)	199(4.8)
Missing	2,316	0	0
Education			
Less than high school	715 (7.8)	5,616 (12.9)	887 (38.0)
High school graduate	2,115 (23.1)	$12,441 \ (28.6)$	1,055 (45.2)
Some college	3,222 (35.2)	11,365 (26.1)	248 (10.6)
College/Some graduate	3,106 (33.9)	14,043 (32.3)	145 (6.2)
Missing	2,156	0	1,775
Self-reported alcohol use			
Positive	3,811 (58.4)	23,904 (55.0)	-
Negative	2,720 (41.6)	$19,561 \ (45.0)$	-
Missing	4,783	0	4,110
Oral alcohol test			
Positive	234 (2.8)	-	-
Negative	8,149 (97.2)	-	-
Missing	2,931	43,465	4,110
Blood alcohol test			
Positive	137(2.7)	-	1,027 (44.7)
Negative	4,926 (97.3)	-	1,268 (55.3)
Missing	6,251	43,465	1,815
Self-reported cannabis use			
Positive	905 (11.7)	3,417 (7.9)	-
Negative	6,813 (88.3)	40,048 (92.1)	-
Missing	3,596	0	4,110
Oral cannabis test			
Positive	864 (10.3)	-	-
Negative	7,519 (89.7)	-	-
Missing	2,931	43,465	4,110
Blood cannabis test			
Positive	555 (11.0)	-	663 (46.8)
Negative	4,494 (89.0)	-	755 (53.2)
Missing	6,265	43,465	2,692

Table 2. Sensitivity and Specificity of Self-Reported Cannabis and Alcohol Use from the NRS Data Samples with Complete Pairs of Self-Reported Use and Blood Test for Both Cannabis and Alcohol, Stratified by Overall, Age, Sex, Race, Education

	Cannabis Use		Alcohol Use			
	n	Sens(%)	Spec(%)	n	Sens(%)	Spec(%)
Overall						
All-Inclusive	$4,\!221$	62	94	3,022	88	47
Age (years)						
16-20	439	66	88	235	100	58
21-34	1,762	63	91	1,441	86	40
35-49	1,039	62	98	744	88	53
50-64	755	44	98	480	91	52
≥ 65	226	100	99	122	100	52
Sex						
Male	2,361	66	94	1,714	90	40
Female	1,860	56	94	1,308	83	57
Race						
White	2,436	66	94	1,760	95	47
Black	931	56	94	670	71	46
Hispanic	434	70	96	301	90	53
Others	428	62	95	291	100	46
Education						
Less than high school	254	55	96	145	100	54
High school graduate	970	63	95	646	82	53
Some college	1,600	60	93	1,175	84	46
College/Some graduate	1,397	67	95	1,056	95	45

Table 3. Baseline Characteristics (Weighted) of Three U.S. National Data Samples Aged 16 Years and Older: Drivers from the 2013-14 NRS, U.S. population from the 2013-14 NSDUH, Homicide Victims from the 2013 NVDRS after Pooling 20 Imputed Samples

Data Source	NRS	NSDUH	NVDRS
	N = 11,314	N = 43,465	N = 4,110
Characteristics	Frequency (%)	Frequency (%)	Frequency (%)
Age (years)			
16-20	1,487 (13.1)	3,817 (8.8)	489 (11.9)
21-34	4,439 (39.2)	10,414 (24.0)	1,813 (44.1)
35-49	2,771 (24.5)	10,751 (24.7)	997 (24.3)
50-64	1,926 (17.0)	10,755 (24.7)	581 (14.1)
≥ 65	691 (6.1)	7,728 (17.8)	230 (5.6)
Sex			
Male	6,566 (58.0)	20,970 (48.2)	3,340 (81.3)
Female	4,748 (42.0)	22,495 (51.8)	770 (18.7)
Race			
White	6,337 (56.0)	28,458 (65.5)	1,229 (29.9)
Black	2,528 (22.3)	5,131 (11.8)	2,306 (56.1)
Hispanic	1,464 (12.9)	6,622 (15.2)	376 (9.1)
Others	985 (8.7)	3,254 (7.5)	199 (4.8)
Education			
Less than high school	1,001 (8.8)	5,616 (12.9)	1,554 (37.8)
High school graduate	2,736 (24.2)	12,441 (28.6)	1,839 (44.8)
Some college	3,785 (33.5)	11,365 (26.1)	448 (10.9)
College/Some graduate	3,792(33.5)	14,043 (32.3)	268 (6.5)
Self-reported alcohol use			
Positive	6,298 (55.7)	23,904 (55.0)	-
Negative	5,016 (44.3)	19,561 (45.0)	-
Oral alcohol test			
Positive	308(2.7)	922 (2.1)	-
Negative	11,006 (97.3)	42,543 (97.9)	-
Blood alcohol test			
Positive	345(3.0)	1,149(2.6)	1,825 (44.4)
Negative	10,969 (97.0)	42,316 (97.4)	2,285(55.6)
Self-reported cannabis use			
Positive	1,303 (11.5)	3,417 (7.9)	-
Negative	10,011 (88.5)	40,048 (92.1)	-
Oral cannabis test	, , ,	, , ,	
Positive	1,112 (9.8)	2,988 (6.9)	-
Negative	10,203 (90.2)	40,477 (93.1)	-
Blood cannabis test	. ,	, ,	
Positive	1,125 (9.9)	3,197(7.4)	1,873 (45.6)
Negative	10,189 (90.1)	40,268 (92.6)	2,237 (54.4)

Table 4. Odds Ratio of Homicide Victimization in the NVDRS Case Group, as Compared with the NRS Control Group or the NSDUH Control Group Using Four Weighted Logistic Multivariable Models: M1 Shows the Estimate of Regression Coefficients Using the Complete Cases of the NVDRS and NRS Data; M2 Pools the 20 Estimates of Regression Coefficients Using 20 Multiply-Imputed NVDRS and NRS Data; M3 Pools the 20 Estimates of Regression Coefficients Using 20 Multiply-Imputed NVDRS and NSDUH Data after Data Fusion; M4 Pools the 400 Estimates of Regression Coefficients Using 200 Bootstrapped and 2 Multiply-Imputed NVDRS and NSDUH Data after Data Fusion

	NVDRS-	+NSDUH	NVDRS+NRS		
	M1: DFBMI n = 47,582	M2: DFMI n = 47,582	M3: CC n = 5,205	M4: MI n = 15,431	
	OR (95%)	OR (95%)	OR (95%)	OR (95%)	
Age (years)					
35-49	Ref	Ref	Ref	Ref	
16-20	1.43 (1.03, 1.82)	$1.37\ (1.06,\ 1.78)$	$1.10 \ (0.67, \ 1.79)$	$0.51\ (0.38,\ 0.69)$	
21-34	1.79 (1.33, 2.25)	$1.84\ (1.47,\ 2.29)$	$1.13 \ (0.79, 1.64)$	$0.81\ (0.68,\ 0.95)$	
50-64	$1.38\ (1.04,\ 1.73)$	1.46 (1.10, 1.94)	$0.83 \ (0.56, 1.24)$	$0.79 \ (0.66, \ 0.96)$	
≥ 65	$0.72\ (0.52,\ 0.92)$	$0.70\ (0.53,\ 0.93)$	$1.06 \ (0.51, \ 2.19)$	$1.48 \ (1.05, \ 2.09)$	
Sex					
Female	Ref	Ref	Ref	Ref	
Male	$2.94\ (2.59,\ 3.3)$	$2.91\ (2.59,\ 3.27)$	2.19 (1.65, 2.91)	1.73 (1.45, 2.07)	
Race					
White	Ref	Ref	Ref	Ref	
Black	5.27(4.55, 6)	5.19 (4.51, 5.97)	2.56 (1.32, 4.96)	$2.61\ (1.60,\ 4.27)$	
Hispanic	$0.73 \ (0.6, \ 0.86)$	$0.72\ (0.59,\ 0.87)$	2.59(1.51, 4.42)	$0.88 \ (0.59, 1.30)$	
Others	$1.6\ (1.29,\ 1.92)$	$1.49 \ (1.15, \ 1.93)$	1.84 (1.03, 3.28)	$1.19 \ (0.78, 1.80)$	
Education					
High school graduate	Ref	Ref	Ref	Ref	
Less than high school	1.68 (1.44, 1.91)	1.69(1.45, 1.97)	$2.23\ (1.68,\ 2.97)$	2.47 (1.97, 3.09)	
Some college	$0.32\ (0.26,\ 0.38)$	$0.31\ (0.26,\ 0.38)$	$0.19 \ (0.14, \ 0.26)$	$0.20\ (0.17,\ 0.25)$	
College graduate/Some graduate	$0.19 \ (0.15, \ 0.23)$	$0.19 \ (0.16, \ 0.24)$	$0.18 \ (0.12, \ 0.27)$	$0.14\ (0.11,\ 0.19)$	
Drug					
Alcohol (positive vs. negative)	19.25 (12.25, 26.24)	19.09 (11.51, 31.65)	15.83 (9.89, 25.33)	17.99 (12.87, 25.15)	
Cannabis (positive vs. negative)	3.55 (2.75, 4.35)	3.59 (2.77, 4.64)	2.58 (1.67, 3.97)	3.53 (2.60, 4.79)	

Distribution of the 10-fold CV Estimated AUC (Cannabis) Distribution of the 10-fold CV Estimated AUC (Alcohol) Middle box line represents the mean of AUC Middle box line represents the mean of AUC Lasso -Lasso -Model Logistic -Logistic -Random_Forest -Random_Forest -0.8 0.9 0.9 1.0 0.7 1.0 0.7 0.8 Area under the ROC curve Area under the ROC curve Distribution of the CV Estimated Sensitivity (Alcohol) Distribution of the CV Estimated Sensitivity (Cannabis) Middle box line represents the mean of sensitivity Middle box line represents the mean of sensitivity Lasso · Lasso -Logistic -Logistic -Random_Forest -Random_Forest -0.00 0.00 1.00 0.25 0.50 1.00 0.25 0.50 0.75 Sensitivity Sensitivity Distribution of the CV Estimated Specificity (Cannabis) Distribution of the CV Estimated Specificity (Alcohol) Middle box line represents the mean of specificity Middle box line represents the mean of specificity Lasso -Lasso -Model Logistic -Logistic -

Random_Forest

0.95

0.96

0.98

Specificity

0.99

1.00

0.98

Specificity

0.99

1.00

Random_Forest

0.95

0.96

Figure 2. Cross-Validated Blood Test Variables Prediction Results