

Reverse Engineering Interstate Crosscheck

By Michael Agosta, VVN Managing Director

Executive Summary

Federal law requires that states periodically perform maintenance on their voter rolls, removing citizens who may have moved or died. Interstate Crosscheck is one such program, designed to cull registered voters who have moved across state lines. The program is touted primarily by conservative politicians who contend that voter fraud in America is rampant, despite considerable research to the contrary. Crosscheck's database methodology is dangerously simplistic and can remove voters who are in fact still residing at their current address. Despite recent setbacks to Crosscheck, the Trump administration remains strong cheerleaders of the program and Crosscheck is explicitly named in the Republican national platform's planks on electoral reform.

This report addresses the checkered history of the Crosscheck system, and examines its potential impacts. VVN performed a full nationwide reverse engineering analysis of Crosscheck. Our simulation found that Crosscheck could potentially expunge over 8% of the names on the national voter roll. We examine the impact of Crosscheck on the Rising American Electorate, paying particular attention to race.

Background: Maintaining Accurate Voter Files

Voter rolls are messy. State and county election officials are not routinely notified when voters move or die, and inaccuracies abound. In 2012, the PEW Center on the States estimated that 24 million voter registration records (comprising roughly 13% of the nation's total registered voters) contained incorrect or outdated information, and an additional 1.8 million deceased individuals remained on the rollsⁱ

A recent FiveThirtyEight story illustrates the problem:

“We did a quick survey of FiveThirtyEight staffers by checking voter registration rolls in the states they've lived in over the past 15 years. Out of 15 people who participated, five were double-registered. I'm one of them, with active voter registrations in Minnesota, where I live, and Alabama, a state I last lived in in 2006. Three staffers were only registered in states they no longer live in. One person wasn't registered *anywhere*, much to his surprise. Bottom line: Americans don't stay in one place forever, and bureaucracy doesn't always keep up with us.”ⁱⁱ

The National Voting Rights Act (NVRA) requires states to periodically purge voter rolls of registrants who have moved out of state, have changed their address within the state, or have

died. Registrars and Secretaries of State may choose from a variety of purge methods, providing the methods chosen are uniform and non-discriminatory. Voter files are cleaned by matching them against a plethora of state and federal datasets. At the federal level, election officials look to the Social Security Administration database, to find deceased registrants, and the U.S. Postal Service's National Change of Address (NCOA) system, to find registrants who have moved. State agency data, such as the Department of Vital Statistics, or the Department of Motor Vehicles (DMV), may also be consulted. More controversially, some states seek to determine the citizenship status of registered voters, consulting the Homeland Security Systematic Alien Verification for Entitlements (SAVE) database. Texas and Kansas look to jury duty lists to discern if potential jurors disqualified on the basis of citizenship appear as registered voters.ⁱⁱⁱ

Interstate Crosscheck, Kris Kobach, and the Election Integrity Commission

In addition to the above, states cooperate to determine out of state moves using two main coalition resources. The Electronic Registration Information Center (ERIC), founded by the PEW Charitable Trusts, and Interstate Crosscheck, founded by Kansas Secretary of State Kris Kobach.

ERIC is highly regarded by database and data security professionals. The program uses a sophisticated algorithm for finding cross state movers and provides advanced data security measures for handling sensitive voter file data. ERIC seeks not only to identify duplicate registrations but further urges participating states to identify and contact eligible unregistered adults to join the voter rolls.

Interstate Crosscheck is a far more controversial operation, and has been roundly pilloried by voter protection and civil rights organizations. Crosscheck was created in 2005 as an alliance between Iowa, Kansas, Missouri, and Nebraska. While Crosscheck is contentious, it is also provided to states free of charge while ERIC carries a per state fee of \$25,000. In an era of tight state election budgets, free sells. Crosscheck grew to cover 30 states in 2016.^{iv}

Crosscheck's data methodology is naïve and produces dubious results. Secretary Kobach seems more intent on "proving" widespread voter fraud than in improving the quality of the voter lists. Crosscheck strives to ferret out and prosecute individuals who have cast ballots in more than one state in the same election ("double voters"). Double voting is a federal felony, but significant research on double voting has been conducted and has uniformly concluded that it is exceedingly rare.^v

Unmoved by the research, Secretary Kobach continued to maintain that voter fraud is widespread. In 2015, Kobach convinced the Kansas legislature to grant him the unprecedented power to directly prosecute election fraud and double voting, proclaiming that he knew of 100 double voting cases in Kansas alone.^{vi}

In July, 2017, Kobach avowed, "We have discovered 128 specific cases of non-citizens who either registered to vote or attempting [sic] to register to vote. But that's just the tip of the iceberg. One expert in the case estimated the total number could be in excess of 18,000 on our [Kansas state] voter rolls."^{vii} Despite the inflated allegations, as of March, 2018, Kobach had garnered just 9 successful prosecutions, a majority of whom were older Caucasian men.

Kobach attempted to take Crosscheck national when he was appointed vice chairman of the newly formed Presidential Advisory Commission on Election Integrity, created by Executive Order on May 11, 2017 by President Trump. From the outset, the Commission was poorly received. By placing Vice President Pence at the head of the Commission with Kobach riding shotgun, administration claims of bipartisanship rang hollow, since its leaders were both Republican. Hyperbolic rhetoric at the first Commission meeting led many voter rights advocates to fear that the Commission was a shill to "prove" Trump's oft-Tweeted assertion that 3 million non-citizens cast ballots in the 2016 general election. The Commission ordered states to submit their voter files for inspection. The request was thwarted by a bipartisan gaggle of Secretaries of State, who either outright refused the Commission's request to submit their state voter rolls for examination or agreed only to a partial submission of the data, citing privacy laws. Failing to garner enough data to proceed, the Commission was disbanded by President Trump on January 3, 2018.^{viii}

Kobach has been the center of a firestorm in the months since the Election Integrity Commission folded, in his vociferous defense of the Kansas Secure and Fair Elections Act (SAFE). The statute requires Kansans to show proof of citizenship, such as a passport or birth certificate, when registering to vote. Since proof of citizenship is not required under the NVRA, the ACLU and the Kansas League of Women Voters brought suit in *Fish v. Kansas*. Kobach has repeated to the court his claim that 18,000 non-citizens are on the Kansas voter registration rolls, but failed to show evidence. The court, under a preliminary injunction, instructed Kobach to reinstate 35,000 voters to the Kansas rolls, who had been denied registration by SAFE. Kobach refused, and in April, 2018 he was found in contempt of court.^{ix}

While Kobach's hijinks in the courts and with the Election Integrity Commission have damaged the reputation of Crosscheck (some 8 states have dropped out of Crosscheck in the past year), it may still be used to "clean" our national voter rolls^x. The Trump administration may turn to the Department of Homeland Security or the Department of Justice in an attempt to obtain state voter rolls where the Election Integrity Commission failed. It is further important to note that a national Crosscheck program was explicitly enumerated in the 2016 Republican National Platform.^{xi}

Crosschecks Data Methodology: Flawed and Simplistic

Crosscheck's data methodology is dangerously simplistic. The program collects voter files from participating Secretaries of States and provides a literal comparison on just three elements: first name, last name, and date of birth. Instances in which the data triad appears on more than one state voter file are flagged as "possible double votes." VVN research determined that the probability of a false positive on a literal match for these three fields is high.

VVN has conducted a comprehensive simulation of the Crosscheck program, using a full national voter file enhanced with an individual level race model, which appends race to every voter in the nation, providing a comprehensive view of the possible impacts of the program. We found a number of surprising results. We hope this analysis will improve understanding of the Crosscheck program, highlighting the facts while dispelling widely-held myths.

If Crosscheck had been applied to the national voter file, as the President's Election Integrity Commission sought, it would have rendered well over 15,669,439 exact first name, last name, and birth date duplicates across the nation, roughly 8% of all registered voters..

The Mechanics of Crosscheck

Some press reports, particularly accounts on social media, incorrectly assert that Crosscheck ONLY uses first name, last name and birthdate to purge voters:

"...election experts have criticized Crosscheck for producing large numbers of matches because the program only cross-checks first names, last names, and dates of birth."^{xii}

"The system identifies voter registrations that have identical first names, last names and dates of birth."^{xiii}

While Crosscheck initially uses a literal comparison of the first name, last name, and date of birth, in fairness the program uses these three fields as a first step. These potential dupes are delivered to the Secretaries of State, advising that the data contain many false positives. A Crosscheck user guide from 2014, acquired by VVN, states that "Experience in the crosscheck program indicates that a significant number of apparent double votes are false positives and not double votes."

The guide further advises participants to take the following next steps:

"An apparent duplicate registration is produced when the first name, last names, and dates of birth in two records match exactly. Other information such as middle name, suffix, and SSN4 [the last four digits of the Social Security number] should be used to confirm whether the two

records are matches. It may be necessary to contact another jurisdiction to obtain more information, such as signatures.”

Crosscheck thus leaves the hardest part of the process (comparing signatures or Social Security numbers, contacting various out of state authorities etc.) up to the discretion of the Secretaries of State. Many states simply do not have the time or staff to follow up, and Crosscheck’s potential double voter lists are thus not ultimately used in voter file purges. Kentucky withdrew from the program, calling the data “unreliable” and Maryland more bluntly asserted the raw data were “a waste of time.”^{xiv}

Surname and First Name Propensity

When analyzing the potential effects of Crosscheck, it is important to keep basic laws of probability in mind. In a subway car with 70 passengers, there is a 99.9% probability that two passengers will share the same birth DAY (month and day). At 180 passengers on the subway car, there is now a 50% probability that 2 will share the exact same birth DATE (day, month, and year). These probabilities illustrate a fundamental flaw in the Crosscheck methodologies: the system will produce a high number of false positives.^{xv}

Many analysts, often examining surname propensity data from the U.S. Census, inadvertently overstate the impact of Crosscheck on voters of color, and the analysis is oft cited in partisan-leaning publications. The report below, from Mark Swedlund, who often provides analysis for reporter Greg Palast and the UK newspaper, The Guardian, typifies the error:

“[Mark] Swedlund's statistical analysis found that African-American, Latino and Asian names predominate, a simple result of the Crosscheck matching process, which spews out little more than a bunch of common names. No surprise: The U.S. Census data shows that minorities are overrepresented in 85 of 100 of the most common last names. If your name is Washington, there's an 89 percent chance you're African-American. If your last name is Hernandez, there's a 94 percent chance you're Hispanic. If your name is Kim, there's a 95 percent chance you're Asian.”^{xvi}

VVN’s analysis on the national voter file rendered the following counts based on racial composition:

Duplicates by Race in VVN’s Crosscheck Simulation			
Race	Total National Voter File	Total Individuals in Crosscheck	% of Cohort in Dupes
Asian	5,398,165	343,748	6.37%
Black	24,322,955	1,967,238	8.09%
Caucasian	143,165,216	12,024,615	8.40%
Hispanic	18,562,627	1,055,025	5.68%

Surprisingly, blacks and Caucasians shared a roughly equal probability of being captured in a potential cross state duplicate. 8% of blacks on the national voter file were captured in a literal match compared with nearly 8.5% of Caucasians. Hispanics and Asians were ensnared less frequently in Crosscheck pairs.

On its face, the surname analysis is correct: blacks have a proportionally smaller number of common surnames. In examining counts for the top 10 most commonly occurring surnames in Crosscheck among blacks and Caucasians, we find that the top 10 black surnames account for nearly 19% of all blacks captured by Crosscheck, but only 9% percent of the Caucasians:

Top 10 Black Surnames in the VVN Simulation	
Last Name	Total in Crosscheck
WILLIAMS	74,011
JOHNSON	55,474
SMITH	48,847
JONES	44,851
BROWN	41,049
JACKSON	28,146
DAVIS	24,816
THOMAS	20,260
HARRIS	18,092
ROBINSON	16,307
TOTAL (18.9%)	371,853

Top 10 Caucasian Surnames in the VVN Simulation	
Last Name	Total in Crosscheck
SMITH	296,184
JOHNSON	163,827
MILLER	123,107
BROWN	109,838
JONES	99,398
WILLIAMS	87,569
DAVIS	81,273
ANDERSON	60,054
WILSON	58,488
TOTAL (9%)	1,079,738

So, why does this analysis fail to explain the Crosscheck results? The answer lies in the distribution and variety among first names.

When first name propensity is included in the analysis, a more complex picture emerges. The true impact of Crosscheck is the probability of two individuals matching on first name, last name, and birth date.

In the case of blacks, first names exhibit an extraordinarily rich variety and a host of variant spellings, and the popularity of any given name waxes and wanes over time. Caucasians have fewer, more frequently occurring first names than blacks. Indeed, blacks flagged as potential duplicates in the Crosscheck methodology had more distinct first names than surnames. The end result: Crosscheck impacts both blacks and Caucasians fairly equally:

Name Frequency by Race In the VVN Crosscheck Simulation		
Race	Total First Names	Total Last Names
Asian	39,842	45,126
Black	115,498	109,057
Caucasian	108,174	474,777
Hispanic	33,995	71,719

In examining the first name data among blacks, we find a rich variety of spellings: Vandra and Vandre; Vandrea and Vandria, etc. “Michael” was the most common first name for both blacks and Caucasians. Among Caucasians, the name appears 321,411 times, accounting for 2.6% of Crosscheck duplicates. Among blacks, “Michael” appears 31,971 times, accounting for 1.6% of the duplicate Crosscheck pairs. Again, first name variance accounts for the levelling out of the impact of Crosscheck among blacks and Caucasians. Blacks have fewer surnames as a cohort, but far more first names.

An MSNBC study found similar results to VVN’s work. In aggregate, Crosscheck does not disproportionately impact people of color. However, the MSNBC analysis found wide variance of Crosscheck results at the local level. Some communities of color, as well as some less affluent white working class neighborhoods, bear more of the brunt than a typical majority Caucasian suburban neighborhood. While the MSNBC study was not nationwide, the findings correlate consistently with the VVN study.^{xvii}

Generally, the impact of Crosscheck is more diffuse, but not necessarily less prevalent, among Caucasians. In Ohio, 634,082 voters were flagged in the national Crosscheck data run. Of that total, 535,431 were Caucasian and 74,522 were black. Blacks are highly concentrated in Franklin County (Columbus), Cuyahoga County (Cleveland), and Hamilton County (Cincinnati) so the impact at the neighborhood level among blacks in urban areas is more pronounced. Caucasians

caught in Crosscheck's crosshairs display less concentration at the neighborhood level, with a more even distribution across the state's 88 counties. In many states, ironically, Crosscheck likely has a greater overall impact on conservative voters. In Ohio, Crosscheck rendered 278,749 possible matches in counties carried by Hillary Clinton, 44% of the total, and 354,236 matches in counties carried by Donald Trump, 56% of the Crosscheck total.

Crosscheck: Checkered Past and an Unsure Future

Even though Crosscheck may not unduly impact voters of color in aggregate, it can be misused. As we have shown, Crosscheck urges election officials to use additional data points, beyond first name, last name, and birth date before purging voters, but the system does not require it. The actual impact of crosscheck in voter file purges is difficult to discern. Election officials are typically opaque or evasive in explaining how Crosscheck data are used in voter file purges, leaving fair election proponents dangerously in the dark.

What is sadly clear is that states are still searching for adequate means to remove movers and deceased voters from their rolls, and Crosscheck, despite its myriad flaws, retains a great many advocates on the political right.

The waters have been further muddied by the Supreme Court ruling on June 11, 2018, in *Husted v. A. Philip Randolph Institute*. In the Case, the state of Ohio removed voters who had not cast a ballot in a federal election over a six year period, jeopardizing many hundreds of thousands of registrations. The Court allowed that while the Ohio method was not ideal, neither was it illegal, setting an insidious precedent. Facing a Court profoundly reluctant to provide specific guidance on the NVRA's voter purge provisions, and facing legislative inaction and gridlock at the federal level, Secretaries of State are left to flounder in a fog of uncertainty. Officials intent on a quest to eliminate non-existent fraud are free to use questionable or incorrect data methodology, more informed by the partisan politics in a given state than the exigencies of solid, fair, or just public policy.

Progressives must remain diligent. Secretary of State Kobach's antics have cast doubt on Crosscheck's future with a bipartisan group of Secretaries of States, but the program remains popular among partisans, and the battle for a truly efficacious system of voter file purges is far from over.

Appendix 1: Federal Legislation and Voting

Three major pieces of legislation provide federal guidance for US Elections:

The Voting Rights Act of 1965: Signed into law by President Lyndon Baines Johnson, the Act explicitly guaranteed the voting rights enumerated in the 14th and 15th Amendments to the US Constitution. The Act expressly prohibited limiting access to the ballot to racial or language minorities and outlawed poll taxes and other voter suppression methods employed predominantly in the South.

The National Voter Registration Act of 1993 (NVRA), Often called the "motor voter bill," because the NVRA required that Department of Motor Vehicle offices enable citizens to register vote while renewing or obtaining a state drivers' license or id card. The NVRA explicitly instructs states to perform routine file maintenance, or purges, on their voter rolls.

The Help America Vote Act of 2002 (HAVA): After the 2000 election debacle, Congress revisited voting rights and enacted the Help America a vote Act of 2002. HAVA dealt with the technical requirements of maintaining voter laws, and ordered the various Secretary of State (SOS) offices to maintain a deduplicated master list of voters for each state. Previous to HAVA, voter lists were primarily maintained at the county level and data quality and duplication issues were epidemic.

ⁱ Pew Trusts. February, 2012.

http://www.pewtrusts.org/~media/legacy/uploadedfiles/pcs_assets/2012/pewupgradingvoterregistrationpdf.pdf

ⁱⁱ FiveThirtyEight, September 22, 2017. <https://fivethirtyeight.com/features/kris-kobach-can-prove-u-s-elections-are-messy-but-thats-not-the-same-thing-as-fraudulent/>

ⁱⁱⁱ National Conference of State Legislators. <http://www.ncsl.org/research/elections-and-campaigns/voter-list-accuracy.aspx>

^{iv} Govtech, November 10, 2017. <http://www.govtech.com/security/Multi-State-Voter-Crosscheck-System-Raises-Security-Questions.html>.

See also the National Council on State Legislatures. <http://www.ncsl.org/research/elections-and-campaigns/voter-list-accuracy.aspx>

^v One Person, One Vote: Estimating the Prevalence of Double Voting in U.S. Presidential Elections. Goel, Meredith et. al. October, 2017.
<https://scholar.harvard.edu/files/morse/files/1p1v.pdf>

^{vi} The Brennan Center for Justice, January 31, 2017.
<https://www.brennancenter.org/analysis/debunking-voter-fraud-myth>

^{vii} The Wichita Eagle, July 19, 2017. <http://www.kansas.com/news/politics-government/article162505228.html>

^{viii} The White House. <https://www.whitehouse.gov/articles/presidential-advisory-commission-election-integrity/> <https://www.nytimes.com/2018/01/03/us/politics/trump-voter-fraud-commission.html8state>

^{ix} CNN, April 19, 2018. <https://www.cnn.com/2018/04/19/politics/kris-kobach-contempt-of-court/index.html>

^x Gizmodo, January 29, 2018. <https://gizmodo.com/eighth-state-quietly-quit-free-anti-voter-fraud-program-1822514538>

^{xi} The Hill, July 19, 2016. <http://thehill.com/blogs/ballot-box/288302-gop-platform-calls-for-tough-voter-id-laws>

^{xii} Courthouse News, October 27, 2017. <https://www.courthousenews.com/indiana-accused-using-flawed-data-purge-voters/>

^{xiii} WBUR Radio, November 3, 2017.
<http://www.wbur.org/radioboston/2017/11/03/massachusetts-crosscheck-system>

^{xiv} Gizmodo, January 29, 2018. <https://gizmodo.com/eighth-state-quietly-quit-free-anti-voter-fraud-program-1822514538>

^{xv} Wikipedia. https://en.wikipedia.org/wiki/Birthday_problem

See also McDonald, Michael P. and Levitt, Justin, Seeing Double Voting: An Extension of the Birthday Problem (July 1, 2007). 7 Election L. J. 111 (2008); 2nd Annual Conference on Empirical Legal Studies Paper. Available at SSRN: <https://ssrn.com/abstract=997888> or <http://dx.doi.org/10.2139/ssrn.997888>

See Also the Washington Post, July 2017.

https://www.washingtonpost.com/news/wonk/wp/2017/07/20/this-anti-voter-fraud-program-gets-it-wrong-over-99-of-the-time-the-gop-wants-to-take-it-nationwide/?utm_term=.0c452412cc8f

^{xvi} Greg Palast, in The Rolling Stone. August 24, 2016.

<https://www.rollingstone.com/politics/features/the-gops-stealth-war-against-voters-w435890>

^{xvii} NBC News, August 24, 2016. <https://www.nbcnews.com/news/us-news/do-voter-purges-discriminate-against-poor-minorities-n636586>