



Analysis of the September 15, 2005 Voter Fraud Report Submitted to the New Jersey Attorney General

by
The Brennan Center for Justice at NYU School of Law
and
Dr. Michael McDonald of George Mason University

December 2005

According to a recent report submitted to the New Jersey Attorney General, 63-year-old K_____ Sullivan was so determined to vote in the November 2, 2004, election that she voted in person in Cape May and then drove to Bergen – or vice versa – traveling 161 miles across the length of the state so that she could vote in person a second time.

There is something bizarre afoot here, but it probably has little to do with Ms. Sullivan. Rather, the problem is with the report – a flawed partisan document stirring up the specter of voter fraud by listing thousands of allegedly illegitimate voters. Among other things, the report included lists of purportedly illegitimate votes in New Jersey in the 2004 general election, including lists of 10,969 individuals who purportedly voted twice and lists of 4,756 voters who were purportedly dead or incarcerated in November 2004. These lists were submitted to the Attorney General in mid-September, as exhibits to a report demanding that New Jersey counties purge their voter files based on the findings therein. The report’s findings were widely publicized, in New Jersey and elsewhere.

We obtained the lists of voters submitted to the New Jersey Attorney General, as well as a copy of the New Jersey county voter registration files, and have conducted an initial investigation of the report’s claims.

Our analysis shows that the lists submitted to the New Jersey Attorney General are substantially flawed, and must not be used to interfere with New Jersey citizens’ right to vote.

These suspect lists were compiled by attempting to match the first name, last name, and birth date of persons on county voter registration files. Entries that supposedly “matched” other entries were apparently deemed to represent the same individual, voting twice. This methodology is similar to the method used in compiling the notoriously inaccurate Florida “purge lists” of suspected ineligible felons in 2000 and 2004. As Florida’s experience shows, matching names and birthdates in the voter registration context is a tricky business, and can easily lead to false conclusions – as was almost certainly the case here.

Care and caution are required in any attempt to accurately match information in one data system with information in another, but special care must be taken when individual rights are at stake to make sure that unreliable matching results are not used to deprive individuals of their fundamental right to vote. Unfortunately, that level of care appears to be absent from the lists submitted to the Attorney General. As a result, neither these lists nor any that are compiled in a similar manner should form the basis of any decision that would impede any voter's registration status or access to the ballot.

Our review reveals several serious problems with the methodology used to compile the suspect lists that compromise the lists' practical value. For example, the data used in the report from one county appears to be particularly suspect and anomalous, and may have substantially skewed the overall results. In addition, middle initials were ignored throughout all counties, so that "J_____ A. Smith" was presumed to be the same person as "J_____ G. Smith."¹ Suffixes were also ignored, so that fathers and sons – like "B_____ Johnson" and "B_____ Johnson, Jr." – were said to be the same person.

More fundamentally, underlying many of the entries on these lists, and similar lists compiled in Florida and elsewhere, is a presumption that two records with the same name and date of birth must represent the same person. As we explain in this analysis, this presumption is not consistent with basic statistical principles. Even when votes appear to have been cast in two different cities under the same name and birth date, statistics show that voter fraud is not necessarily to blame. With 3.6 million persons who voted in the 2004 election in New Jersey, the chance that some have the same name and birth date is not far-fetched. Which is more reasonable: that there are two K_____ Sullivans born on July 5, or that one 63-year-old woman drove three hours across the state, from Cape May to Bergen, to make sure that she voted twice on Election Day?²

Individual examples like Ms. Sullivan's purported Election Day journey may seem trivial once they are caught and debunked. But the potential for harm is no laughing matter. Using flawed lists like these to purge the registration rolls would inevitably lead to disfranchisement of eligible voters and disruption at the polls when the eligible citizens arrive to vote in the next election. The voters deserve better. New Jersey should forswear use of the lists – and any list constructed with similarly flawed methods – to question the eligibility of voters of any political affiliation.

¹ The names used in this report represent actual entries in datafiles submitted to the New Jersey Attorney General. In order to understand the methodological flaws involved in the compilation of the datafiles, it is important to cite particular examples. But in order to protect individual privacy, we have reprinted only the first initial of each voter's given name. A discussion of "John A. Doe" and "John B. Doe" therefore appears in this analysis as "J_____ A. Doe" and "J_____ B. Doe." In each instance comparing one voter to another, these abbreviations ("J_____" and "J_____") represent the same first name.

² This is, obviously, a rhetorical question highlighting a methodological flaw in the report. But the question also has a real answer: a quick telephone inquiry confirmed that two different New Jersey residents named K_____ Sullivan were born on the same date; each voted separately (and quite legally) in 2004.

Background: The September 15, 2005 Report and Accompanying Lists

The following analysis documents several of the methodological flaws in the lists of voters included as exhibits to the voter fraud report submitted to the New Jersey Attorney General on September 15, 2005. The analysis is based on methodology only: we have not gained access to original documents related to registration or original pollbook records; we have only recently acquired and compiled copies of the counties' original registration datafiles, which contain some notable gaps; and the lists submitted to the Attorney General contain significant errors and little documentation, which complicates the analysis. Nonetheless, the information we have collected so far is sufficient for the purposes of generally assessing the quality of the evidence presented to support the September 15 report. Our review of the suspect lists reveals that the evidence submitted does not show what it purports to show: cause for concern that there is serious risk of widespread fraud given the state of the New Jersey voter registration rolls.

The suspect lists were apparently prepared by assembling and compiling voter registration files from New Jersey's counties. It is clear that these registration files were subjected to data processing to format the data uniformly, but no documentation was submitted with the files to explain how the processing was done. All that we have is the end product: several lists of New Jersey voters that allegedly reveal fraud or the potential for fraud.

Most of these lists were apparently constructed by attempting to find each citizen's registration and voting records – across counties and supplemented by external data – by matching the first name, last name, and date of birth associated with each record. For example, all registration and voting data for any “J_____ Smith” born on June 6, 1960 were apparently compiled and merged into one record, showing that such a J_____ Smith registered in one New Jersey town and voted there in 2004, and that such a J_____ Smith registered in a different New Jersey town and voted there in 2004.

The report submitted to the Attorney General analyzes these compiled records. It concludes – for example – that the J_____ Smith records above show not two “J_____ Smiths” born on the same day, but that a single individual voted twice. Indeed, the report claims, in similar fashion, that the data reveal widespread voter fraud in the 2004 general election, including approximately 4,397 individuals in New Jersey who allegedly voted twice. **These claims, however, are not supported by the evidence submitted.**

Analysis of the Double-Voting Claim

Even a cursory examination of the lists suffices to show that the claims of widespread fraud are greatly exaggerated. For example, although the allegations of multiple voting are presented as though they are based on reliable evidence that the same person has voted more than once, 80% of the relevant records – 3,532 voters on the list – are labeled on the New Jersey double-voting list as “less likely” indications of a double-vote. This caveat appears nowhere in the accompanying report to the Attorney General or in the press reports of the report’s findings.

It is also curious that only 0.6% of the voters on the lists are alleged to have voted more than twice. Why would a miscreant committed to perpetrating voter fraud stop at just two votes, rather than voting again and again and again?

The simplest answer, of course, is that the lists do not show fraud at all, but reveal instead only methodological problems with the way in which they were constructed. Attempts to match data on one list to data on another list will often yield “false positives”: two records that at first appear to be a match but do not actually represent the same person. The natural incidence of “false positives” for a matching exercise of this scale – especially when, as here, conducted with relatively little attention to detail – readily explains the ostensible number of double votes. It also explains the extremely low incidence of individuals who appear to vote three or more times: it is relatively rare that three “false positives” will all be tied to each other, unless the data to be matched is extremely common. And that it is why it is also no surprise that among the 27 citizens labeled on the suspect lists as voting three or more times in 2004, we find individuals with extremely common names, like “P_____ Smith” or “R_____ Miller,” or “L_____ Wong.”

Common sense thus indicates that whatever is going on here is not a sudden wave of voter fraud. Digging a little deeper into the substance of the lists submitted to the New Jersey Attorney General yields even more reason to be skeptical of the double-voting claims. Consider, for example, a detailed analysis of the 4,397 individuals alleged to have voted multiple times in New Jersey in 2004.

1. *Probable data duplication error: same name, same birth date, same address*

1,803 of these 4,397 records of ostensibly illegal votes seem to be the product of a glitch in the compilation of the registration files. These records reflect two registration entries by the same person from the same address, with a notation next to each that the individual has voted. For example, 55-year-old W_____ A. Connors, living at 253 B_____ Ave. in a New York commuter suburb, is listed on the datafiles submitted to the Attorney General with an (erroneous) first registration date in 1901 and a second registration date in 1993; Mr. Connors is thus represented twice on the datafiles submitted. Each of these entries also indicates that W_____ A. Connors at 253 B_____ Ave voted in 2004.

There is no credible indication, however, that Mr. Connors actually *voted* twice; indeed, given the clearly erroneous registration date on the files, it is far more likely that data error is to blame for the doubly logged vote as well. To believe that Mr. Connors voted twice, we would have to believe that he walked into the polling place, signed the poll book in front of a pollworker, and voted – and then returned to the *same* polling place, again signed the poll book in front of the pollworker, in a spot precisely adjacent to the prior signature, and again voted – all without arousing suspicion. And we would have to believe that 1,802 other voters did precisely the same. We would have expected at least one poll worker to notice one of these 1,803 individuals attempting to sign a poll book under the same name at the same address in the same precinct more than once.³

There is, of course, a more plausible explanation. The bulk of these 1,803 records may be traced to irregularities in the data processing and compilation process for one single county: the Middlesex County registration file accounts for only 10% of registered voters in the state but 78% of these alleged double votes. This does not prove a massive double-voting scheme in Middlesex County. Rather, it indicates a data processing problem with the Middlesex County datafile: a far more likely explanation than voter fraud for the fact that 89% of the Middlesex voters on the suspect lists are listed as double-voting from the same address. Indeed, the suspect lists themselves contain an acknowledgment that the problem in Middlesex is probably not fraud: 99% of these Middlesex voters are labeled on the lists submitted to the Attorney General with a notation that the record is “less likely” to indicate an illegal double vote .

³ Eight of these records reflect registration addresses that are ostensibly different, but which on closer examination almost certainly reflect two records for the same person at the same physical address. Examples include records for individuals with the same name and birth date at “1/2 413 Summer St. 1” and “4131/2 Summer St. 1”; “64 W. 10th St. 2” and “64 W. 10th St.”; and “602 Sooy Ln.” and “604 Sooy Ln.”

2. *Probable data entry errors: different name, same birth date, same address*

Another 1,257 entries of the 4,397 records probably represent similar data errors – also largely driven by a likely glitch in the Middlesex County file, which is also vastly overrepresented in this category. These records show ever-so-slight variations in records listed with the *same date of birth at the same address*: for example, the same first and last names, but different middle initials or suffixes (e.g., J_____ T. Kearns, Sr., and J_____ T. Kearns, Jr., both born the same day and living at the same address; or J_____ E. Allen and J_____ P. Allen, born the same day and living at the same address).

As above, it is extremely unlikely that two votes were actually cast here. Again, the files show errors: J_____ E. Allen is listed as registered in 1901, while J_____ P. Allen, born on the same day and living at the same address, is listed as registered in 2003. And again, to believe that Mr. Allen voted twice, it would be necessary to believe that he arrived twice at the polling place to sign his name in front of a pollworker in adjacent spaces on the poll books, without arousing any suspicion – and that 1,256 other voters did likewise.

Instead, it is far more likely that the listed votes represent a data error. Without access to the original registration records, it is impossible to determine where the mistake lies – for example, whether J_____ T. Kearns, Sr. and J_____ T. Kearns, Jr. were in fact born on different days, or whether the “Jr.” suffix is instead mistaken. A clue to the origin of the problem is that the original Middlesex County voter registration file does not list a birth date for any individual; we do not know how the authors of the September 15 report obtained birth date information for their matching purposes, or what errors may have been introduced during this process. What *is* clear is that the files do not establish a presumption that J_____ T. Kearns, Sr. and J_____ T. Kearns, Jr. represent one individual who fraudulently voted twice.⁴

⁴ Some records in the file submitted to the Attorney General contain three different entries of registration information, so that registration information for F_____ DeLuca, F_____ M. DeLuca, and F_____ C. DeLuca may all be conflated in one record. It is therefore possible for one record to reflect more than one type of the errors or irregularities discussed above.

3. Probable different individuals: different name, same birth date, different address

Approximately 800 of the entries on the list likely represent different people, with different addresses and different middle initials or suffixes. For example, W_____ S. Smith, living in a northern New Jersey town, and W_____ C. Smith, living in another town two hours away, share the same date of birth but are not the same person. Nor are T_____ Brown, living in a New York commuter suburb, and T_____ H. Brown, Jr., living in a small town over an hour west, despite the fact that they also share the same birth date.

About three-quarters of the entries in this category reveal data that affirmatively conflict – for example, a middle initial (“W_____ S.”) in one case, and a different middle initial (“W_____ C.”) in another, listed at different addresses. There is absolutely no good reason to conclude that these individuals are in fact the same, when the available evidence indicates the contrary.

For approximately 200 of the entries in this category, however, less information is available. These entries show a middle initial (“J_____ W. Davis”) in one case, and no middle initial (“J_____ Davis”) in another – again, at different addresses. The lack of the middle initial is ambiguous: it could mean that one of the J_____ Davises in question has no middle name, or it could mean that the middle initial was simply omitted in a particular registration entry. Although these entries involve less conclusive affirmative evidence of a false match than the entries noted above, there is still no good reason to believe that “J_____ W. Davis” and “J_____ Davis,” at different addresses, represent the same person. As explained in further detail below, the incidence of individuals sharing the same name and birth date is sufficiently common that no valid conclusion of a fraudulent vote can be drawn here.

4. Unwarranted assumptions given errors in the listed date of birth

Of the individuals remaining, there are serious concerns with the accuracy of the dates of birth. Seven voters were apparently born in January 1, 1880 – which is most likely a system default for registrations lacking date-of-birth information. For 227 voters, only the month and year of birth are listed: this means only that two voters with the same name were born in the same month and year, an unsurprising coincidence in a state of several million people.

5. *Unwarranted assumptions given the statistics of name-and-birthday matching*

That leaves approximately 289 votes cast under the same name and birth date – like votes cast by “P_____ S. Rosen,” born in the middle of the baby boom – but from two different addresses.⁵ It may appear strange, but there may be two P_____ S. Rosens, born on the same date in 1948 – and such coincidences are surprisingly common. Of course, for any one person, the odds of someone else having the same name and birth date is small. But because there are so many voters in New Jersey, a sizable number will have the same name and birth date simply by chance.

In a modestly sized group, the probability that two people have the same birthday – day and month – is, for many observers, surprisingly high. In a group of just 23 people, it is more likely than not that two will share the same birthday. For 40 people, the probability is 90%. These statistics are known as the “Birthday Problem,” which is often used to introduce students to counterintuitive results in statistics and probability.

Applying the “Birthday Problem” to voter registration lists is fairly straightforward. By including the year (and thus the full birth date), the statistics change somewhat, but the threshold is still surprisingly small to many: given some reasonable assumptions about the average lifespan, the probability that at least two of 150 people have the same exact birth date – day, month, and year – is 50%. And in a group of 300 people, the probability that two share a birth date match is approximately 90%.

Imagine that our group contains all of the registered New Jersey voters with a given first name and last name – such as all of the 417 Robert Smiths who are listed on New Jersey records as voting in 2004. The probability that at least two of these 417 individuals have the same birth date – day, month, and year – approaches 100%. The fact that two Robert Smiths with the same birthday voted in 2004 thus indicates not voter fraud, but a straightforward application of the “Birthday Problem.”

Indeed, the probabilities above likely *underestimate* the chance that a group of voting Robert Smiths share the same birth date, because the above calculations assume that birthdays are randomly distributed when, in fact, they are not. Certain given names are more common in certain years (it would be unsurprising to find two Jessica Smiths born on the same day in 1985, or Lisa Smiths in 1965, or Mildred Smiths in 1925). Likewise, the prevalence of surnames will fluctuate with the immigration patterns of particular ethnicities, which vary from decade to decade. Because older individuals vote at higher rates than younger people, too, we would expect a clustering of voting “Robert Smiths” weighted toward years past. Finally, birth dates themselves are not evenly distributed, as obstetricians are more likely to induce labor during the work week.

⁵ Actually, we know only that these records share a first name, last name, and middle initial – but we cannot know from the lists submitted whether the same middle initial represents the same middle name. The “S.,” for example, could stand for “Stephen” or “Samuel” or “Stanley” or “Stuart,” or any other name beginning with “S,” or – famously following the sitting President – “S” could stand for nothing at all.

First Name	Last Name	Frequency
ROBERT	SMITH	417
JAMES	SMITH	368
WILLIAM	SMITH	365
JOHN	SMITH	321
MICHAEL	SMITH	256
ROBERT	JOHNSON	252
MARY	SMITH	249
ROBERT	BROWN	247
MARIA	RODRIGUEZ	242
ROBERT	MILLER	242
JOSE	RODRIGUEZ	241
THOMAS	SMITH	234
JAMES	BROWN	227
JOHN	MILLER	226
RICHARD	SMITH	223
JOHN	MURPHY	216
DAVID	SMITH	216
ROBERT	WILLIAMS	211
PATRICIA	SMITH	207
BARBARA	SMITH	204
JAMES	JOHNSON	204
WILLIAM	JOHNSON	202
JOHN	WILLIAMS	202
ROBERT	JONES	199
JOHN	KELLY	196

**25 Most Common Names of 2004 Voters
from New Jersey Voter Registration Files**

To demonstrate the magnitude of potential birth date matches, we provide the 25 most common first and last names of persons with a record of voting in the 2004 general election, as recorded on the New Jersey voter registration rolls. For every one of these combinations, it is statistically more likely than not that at least two voters share the same exact birth date. Even for less common names, a shared birth date is less rare than one would intuit – and with almost 150,000 names listed at least twice on the voting rolls, it is not surprising to discover that more than a handful share the same birth date. Many, if not most, of the 289 alleged double votes of persons registered at different addresses most likely reflect two separate individuals sharing a first name, last name, middle initial, and birth date.⁶ The suspect lists alone do not establish convincing, much less conclusive, evidence of widespread fraud.

⁶ Triple birth date matches are rare, but not impossible, given the frequency with which some names appear on the rolls. It is therefore telling that only 0.6% of the alleged fraudulent votes – rare, but not impossible – are attributed to more than three or more votes under one name.

Implications for the Other Claims in the Sept. 15 Report

Similarly, the other claims asserted in the report to the Attorney General – that New Jersey votes cast in the 2004 election reflect the same name and birth date as deceased and incarcerated persons and people in other states – are more likely a statistical coincidence or the result of data errors than evidence of actual vote fraud. And there are methodological problems with the report’s other claims beyond the issues addressed above. For example, the report’s curious use of the country’s *most* common names in attempting to match incarcerated individuals to 2004 votes actually *increases* the chance that the same name and the same birth date will be shared by two different individuals. We cannot fully evaluate these claims at this time since we do not currently have access to the lists of deceased individuals, prisoners, and voters in other states that were used as grist for the claims submitted, but if such lists or files are made available to us, we will review the claims thoroughly to determine their merit.

Finally, the September 15 report makes much of the raw potential for foul play based on the unsurprising fact that there are voters who appear on the New Jersey registration rolls more than once. As noted above, many of the names identified reflect two different individuals and not simply duplicate entries. But we have no doubt that there are duplicate entries on New Jersey’s registration rolls. It is well-known that voter registration rolls contain “deadwood” – registration entries for individuals no longer living at a given address or deceased. There is no evidence, however, that these extra registrations are used for widespread illegal voting. Moreover, the problem of deadwood will soon be largely resolved: both the National Voter Registration Act of 1993 and the Help America Vote Act of 2002 require states to implement several systems and procedures as of January 1, 2006, that will clean the voter rolls of duplicate or invalid entries while protecting eligible voters from unintended disfranchisement.

Unlike the voter list maintenance systems envisioned by Congress, using lists compiled like the flawed lists here as a basis for purging the voter rolls affords voters no protections from erroneous disfranchisement. In their haphazard assessment of ostensible fraud, these lists are massively overinclusive, and if used improperly, are virtually guaranteed to exclude many eligible voters. The purges and challenges nearly implemented in Florida and Ohio in 2004 revealed the dangers of list management on the fly; we should learn from those mistakes.

Conclusion

It is certainly possible that of the 3.6 million New Jersey citizens voting in 2004, a handful managed to vote twice. Such cases can and should be investigated under existing law, and some of the information submitted to the Attorney General might prove to be helpful for that purpose. But the September 15 report and its accompanying lists of purportedly illegitimate voters simply fail to support the overall charge: an incidence of voter fraud sufficient either to cast doubt on the 2004 election or to provoke concern in elections to come. The vast majority of alleged irregularities identified in the report to the Attorney General and its accompanying lists do not stand up to even cursory analysis. One list, for example, was submitted with 80% of the entries specifically labeled in the datafile as “less likely” indicators of illegal activity. Middle names and suffixes were ignored; basic data compilation errors abound. And the report nowhere accounts for the basic statistical reality of the “Birthday Problem.”

The flaws in the data submitted to the Attorney General should preclude use of these lists – or the use of any similarly compiled file or list – as the basis for any purges of the voter files. Procedures established by HAVA and the NVRA, among others, when implemented with the safeguards contemplated by those statutes, will clean the registration rolls without the risk to eligible voters created by suspect lists like those submitted in the September 15 report. After a preliminary review of the evidence submitted, these lists simply do not prove what they purport to prove.

Analysis of Alleged Double-Votes Reveals No Proof of Widespread Double Voting

