IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

STATE OF TEXAS, *Plaintiff*, v. ERIC H. HOLDER, JR., in his Official capacity as Attorney General of the United States, *Defendant*.	Case No. 1:12-cv-00128 (DST, RMC, RLW)

NOTICE OF CORRECTIONS TO THE SUPPLEMENTAL DECLARATION OF DR. THOMAS SAGER

The State of Texas hereby files a revised version of the Supplemental Expert Declaration of Thomas Sager (dated June 11, 2012). The revised version, attached hereto as Exhibit 1, contains corrections identified during Dr. Sager's deposition. The Corrected Supplemental Expert Declaration of Thomas Sager is identical to his original supplemental declaration except for the deposition corrections. The State of Texas therefore withdraws the Supplemental Expert Declaration of Thomas Sager (dated June 11, 2012) from its Exhibit List and replaces that report with the Corrected Supplemental Expert Declaration of Thomas Sager (Dated July 10, 2012). The Corrected Supplemental Expert Declaration will be listed as Exhibit number PX_013 on Plaintiff's Exhibit List.

Dated: July 12, 2012

Respectfully submitted.

GREG ABBOTT
Attorney General of Texas

DANIEL T. HODGE
First Assistant Attorney General

JONATHAN F. MITCHELL
Solicitor General

/s/ *Adam K. Mortara*
ADAM K. MORTARA
JOHN M. HUGHES
ASHA L.I. SPENCER
Bartlit Beck Herman Palenchar & Scott LLP
54 W. Hubbard Street, Suite 300
Chicago, IL 60654
Tel: (312) 494-4400
Fax: (312) 494-4440

PATRICK K. SWEETEN
Assistant Attorney General
ADAM W. ASTON
Principal Deputy Solicitor General
ARTHUR C. D'ANDREA
Assistant Solicitor General
MATTHEW H. FREDERICK
Assistant Attorney General

209 West 14th Street
P.O. Box 12548
Austin, Texas 70711-2548
(512) 936-1695

Counsel for the State of Texas

CERTIFICATE OF SERVICE

I hereby certify that a true and correct copy of the foregoing document is being served by CM/ECF and/or electronic mail on July 12, 2012 on the following:

Elizabeth Stewart Westfall
Jennifer Lynn Maranzano
Daniel J. Freeman
Bruce I. Gear
Meredith E.B. Bell-Platts
U.S. DEPARTMENT OF JUSTICE
Civil Rights Division, Voting Section
950 Pennsylvania Avenue, NW
NWB-Room 7202
Washington, DC 20530
(202) 305-7766/Fax: (202) 307-3961
Email: elizabeth.westfall@usdoj.gov: Email: daniel.freeman@usdoj.gov
Email: bruce.gear@usdoj.gov: Email: jennifer.maranzano@usdoj.gov
Email: Meredith.Bell-Platts@usdoj.gov

Counsel for the United States

Chad W. Dunn
BRAZIL & DUNN
4201 FM 1960 West, Suite 530
Houston, TX 77068
(281) 580-6310
Email: chad@brazilanddunn.com

J. Gerald Hebert
Attorney at Law
191 Somerville Street, #405
Alexandria, VA 22304
Telephone: 703-628-4673
hebert@voterlaw.com

Counsel for Eric Kennie, Anna Burns, Michael Montez, Penny Pope, Marc Veasey, Jane Hamilton, David De La Fuente, Lorraine Birabil, Daniel Clayton, and Sergio Deleon

Ezra D. Rosenberg
Michelle Hart Yeary
DECHERT LLP
902 Carnegie Center, Suite 500
Princeton, NJ 08540
(609) 955-3200/Fax: (609) 955-3259
Email: ezra.rosenberg@dechert.com: Email: michelle.yeary@dechert.com

Jon M. Greenbaum
Mark A. Posner
Robert A. Kengle
LAWYERS' COMMITTEE FOR CIVIL RIGHTS UNDER LAW
1401 New York Avenue, NW, Suite 400
Washington, DC 20005
(202) 662-8325
Email: jgreenbaum@lawyerscommittee.org: Email: mposner@lawyerscommittee.org
Email: bkengle@lawyerscommittee.org

Myrna Perez
Ian Arthur Vandewalker
Wendy Robin Weiser
THE BRENNAN CENTER FOR JUSTICE AT NYU LAW SCHOOL
161 Avenue of the Americas, Floor 12
New York, NY 10013-1205
(646) 292-8329/Fax: (212)463-7308
Email: myrna.perez@nyu.edu: Email: ian.vandewalker@nyu.edu
Email: wendy.weiser@nyu.edu

Robert Stephen Notzon
1507 Nueces Street
Austin, TX 78701-1501
(512) 474-7563
Fax: (512) 852-4788
Email: robert@notzonlaw.com

Victor L. Goode
NAACP National Headquarters
4805 Mt. Hope Dr.
Baltimore, Maryland 21215-3297
(410) 580-5120
Email: vgoode@naacpnet.org

Jose Garza
7414 Robin Rest Dr.
San Antonio, Texas 98209
(210) 392-2856 (phone)
Email: garzpalm@aol.com

Gary L Bledsoe
Law Office of Gary L. Bledsoe and Associates
316 West 12th Street, Suite 307
Austin, Texas 78701
(512) 322-9992
Email: garybledsoe@sbcglobal.net

Counsel for Texas State Conference of NAACP Branches, Mexican American Legislative Caucus of the Texas House of Representatives

Douglas H. Flaum, Michael B. de Leeuw
Adam Harris
FRIED, FRANK, HARRIS, SHRIVER & JACOBSON LLP
One New York Plaza
New York, New York 10004-1980
(212) 859-8000
Email: douglas.flaum@friedfrank.com; Email: michael.deleeuw@friedfrank.com
Email: adam.harris@friedfrank.com

Ryan Haygood, Natasha M. Korgaonkar
Leah C. Aden, Dale E. Ho
Debo P. Adegbile, Elise C. Boddie
NAACP Legal Defense and Educational Fund, Inc.
99 Hudson Street, Suite 1600
New York, New York 10013
(212) 965-2200 / (212) 226-7592
Email: rhaygood@naacpldf.org; Email: nkorgaonkar@naacpldf.org
Email: laden@naacpldf.org; Email: dadegbile@naacpldf.org
Email: dho@naacpldf.org; Email: eboddie@naacpldf.org

Counsel for Texas League of Young Voters Education Fund, Imani Clark, KiEssence Culbreath, Demariano Hill, Felicia Johnson, Dominique Monday, and Brianna Williams

John Kent Tanner
3743 Military Road, NW
Washington, DC 20015
(202) 503-7696
Email: john.k.tanner@gmail.com:

Nancy Abudu, Katie O'Connor
Laughlin McDonald
AMERICAN CIVIL LIBERTIES UNION FOUNDATION INC
230 Peachtree Street NW, Suite 1440
Atlanta, GA 30303
(404) 523-2721
Email: nabudu@aclu.org: Email: koconnor@aclu.org
Email: lmcdonald@aclu.org

Arthur B. Spitzer
AMERICAN CIVIL LIBERTIES UNION OF THE NATION'S CAPITAL
4301 Connecticut Avenue, N.W., Suite 434
Washington, D.C. 20008
(202) 457-0800
Email: artspitzer@gmail.com

Lisa Graybill, Rebecca Robertson
AMERICAN CIVIL LIBERTIES UNION FOUNDATION OF TEXAS
1500 McGowan Street
Houston, Texas 77004
(713) 942-8146
Email: lgraybill@aclutx.org: Email: rrobertson@aclutx.org

Penda Hair, Kumiki Gibson
ADVANCEMENT PROJECT
1220 L Street, NW, Suite 850
Washington, DC 20005
(202) 728-9557
Email: phair@advancementproject.org: Email: kgibson@advancementproject.org

Counsel for Justice Seekers, League of Women Voters of Texas, Texas Legislature Black Caucus, Donald Wright, Peter Johnson, Ronald Wright, Southwest Workers Union and La Union Del Pueblo Entero

Nina Perales
Amy Pedersen
MEXICAN AMERICAN LEGAL DEFENSE & EDUCATIONAL FUND, INC.
110 Broadway, Suite 300
San Antonio, TX 78205
(210) 224-5476
Email: nperales@maldef.org
Email: apedersen@maldef.org

Counsel for Mi Familia Vota Education Fund, Southwest Voter Registration Education Project, Nicole Rodriguez, Victoria Rodriguez

/s/ *Adam K. Mortara*
ADAM K. MORTARA

Exhibit 1

THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

STATE OF TEXAS, *Plaintiff* vs. ERIC H. HOLDER, JR., in His Official Capacity as Attorney General of the United States, *Defendant*.	Case No. 1:12-CV-00128 (RMC, DST, RLW) Three-Judge Court

CORRECTED
SUPPLEMENTAL EXPERT DECLARATION OF THOMAS SAGER

1. I have reviewed the report of the Department of Justice's expert, Dr. Stephen Ansolabehere ("SA").

2. The State has requested that I provide additional analysis of the work of SA. In general my critique of SA's work is limited to his "Protocol for Matching Databases" that appears at ¶¶19-29 of his report. Several of SA's opinions involve qualitative assessments about who is or is not more likely to be affected by SB 14. I note those qualitative assessments, but it is not within the scope of my first declaration nor this supplemental declaration to opine as to their validity.

3. Specifically, I was asked to undertake two additional tasks: (1) I was asked to match SA's VRNID dataset to the State Driver License ("DL") and License to Carry ("LTC") datasets using alternative matching

criteria, as well as to screen for "age over 65" and "voter registration suspense" statuses. Both the State and SA derived lists of registered voters that lack apparent matches to the DL and LTC datasets after application of their matching criteria. In my earlier declaration, I discussed the results of my attempts to find additional matches for the State's list ("May No Match"). I am now asked to attempt to find additional matches for SA's list ("VRNID"). (2) I was asked to consider if there might be statistical and/or data processing reasons that might account for the much larger size of SA's no-match list (VRNID, with 1,501,977 ultimate entries) compared with the State's May No Match list (with 588,095 entries).

4. I understand that SA cleaned both the State voter registration ("VR") and DL/LTC databases before attempting to match them. SA reports that he found 13,072,454 records in the original VR database and that he removed 273 of these records because of duplicate voter ID numbers, 25 because of duplicate SSNs and birth dates, and 6,652 because of duplicate first name, last name, date of birth and street address (SA ¶ 19). SA also reports that he found 125,015 VR records lacking SSNs and having very common names and that he removed these records for most of his analysis (SA ¶ 22).[1]

5. SA reports that he found 25,985,555 records in the original DL database and that he removed numerous records for the reasons listed in the

[1] In SA ¶26, SA refers to these 125,015 records as being in the DL database. I assume he means they are in the VR database.

following table (SA ¶ 20). This left 19,951,173 records in the DL database.[2] He also reports that there are 266,151 DL records with "ambiguous status" that he both included and excluded in two separate runs of his analysis (SA ¶ 22).

SA Deletions from State DL Database Prior to his Matching Sweeps

A	287,236	no drivers license or State ID
B	3,144,900	drivers license expired for more than 2 years
C	1,535,504	drivers license expired between 60 days and 2 years
D	779,918	deceased driver
E	724,974	duplicate SSN
	6,472,532	**TOTAL**

6. SA also reports that he found 840,664 records in the original LTC database and that he removed numerous records for the reasons listed in the following table (SA ¶23). This left 592,270 records in the LTC database.[3]

SA Deletions from State LTC Database Prior to his Matching Sweeps

2,338	Deceased
1	Unreadable
38,919	Non-U.S. citizen
12,437	failed application
194,669	nonrenewable expired license
248,364	**TOTAL**

[2] Apparently, the deletions described by SA in SA ¶ 20 are not mutually exclusive, for the total deletions that he enumerates would leave 25,985,555 – 6,472,532 = 19,513,023 records instead of the 19,951,173 that he asserts.

[3] Subtraction yields 840,664 – 248,364 = 592,300 remaining records – a small discrepancy from the 592,270 reported by SA.

7. With the thus-cleaned databases, SA conducted three consecutive and apparently cumulative sweeps to find matches for VR records in the DL and LTC databases based on the following matching criteria, in the order given:

(Sweep #1) identical SSN

(Sweep #2) identical date-of-birth and identical first name and identical last name (all three criteria required for a match)

(Sweep #3) identical date-of-birth and identical first name and identical last name and identical middle name (all four criteria apparently required for a match)

8. As a side note, it is not clear whether SA's sweep #3 will pick up any additional records. Any records in VR that match DL/LTC on the basis of DOB+FN+LN+MN in sweep #3 will have already been matched on the basis of DOB+FN+LN in sweep #2.

9. As a result of application of his three sweeps, SA matched all but 1,893,143 VR records to DL/LTC (SA ¶ 26). This count includes 125,015 with "insufficient information to match" and 266,151 with "ambiguous DL status" (SA ¶ 26). These 1,893,143 "no match" VR cases constitute SA's VRNID list. It is this VRNID database (reduced by exclusion of the 125,015 and 266,151 aforementioned records) that the State has asked me to try to match to the DL/LTC databases.

10. SA has provided his VRNID database to the State (absent sensitive SSN data). The State was able to resupply the missing SSN information with high confidence by matching on the unique voter ID field. Therefore, the VRNID was available to me for various types of matching strategies. After matching VRNID to DL/LTC, I found that most of the records in VRNID either match DL/LTC or fall into special categories like "over age 65" that I understand are significant for the purpose of SB 14. [4]

11. Before I discuss my attempts to match SA's VRNID to DL/LTC, I will discuss a number of obvious problems with the data cleaning decisions that SA made that have the effect of inflating the number of entries in his VRNID.

12. First, SA removed 779,918 deceased drivers from DL prior to his matching sweeps, but he did not remove them from VR (see category D in my ¶5). Therefore, any deceased drivers who are in the VR database will remain unmatched and will end up in VRNID. Because of this, there are 57,718 deceased persons in SA's VRNID. Presumably, the dead do not vote and therefore should not be included on a list of those potentially affected by SB 14. SA could have avoided this problem by leaving deceased drivers in the

---

[4] Because I was out of the state during the period of this analysis and because of the sensitivity of social security information, the analysis was performed by a technician at the Office of the Attorney General Legal Technical Support division under my remote supervision and direction. I reviewed and quality checked all of the matches by receiving random samples and reviewing them for accuracy.

DL database. By that means, deceased (and formerly driving) voters could have been matched and removed and not end up in VRNID.

13. Second, SA removed 4,680,404 records from DL prior to his matching sweeps on account of expired drivers licenses (categories B and C in my ¶ 5). These removals are problematic for a number of reasons:

14. (i) Some expired licenses no doubt belong to voters who have moved out of State and therefore are no longer eligible to vote in Texas. Such (former) voters will end up in VRNID, although they – like deceased voters – presumably should not be counted among the unmatched.

15. (ii) I understand that registered Texas voters who are over age 65 are automatically entitled to vote by mail and disabled voters are exempt from photo ID requirements. I understand that these exemptions apply regardless of the expiration status of a voter's driver license. No doubt many voters holding expired licenses are over 65 or disabled. Deletion of expired drivers licenses from DL would be unexceptionable for the purpose of age-over-65 determination, provided an additional screen of the VR database for age were added.[5] Disabled voters with expired licenses would end up in VRNID without the capacity to detect their exemption from SB 14 on account of disability.

16. (iii) I understand that drivers holding licenses that are expired less than two years may conveniently renew online, and many in this group

[5] SA did not screen for either disability or age over 65.

may choose to do so. It is clear that the effect of the three factors listed above is to inflate the size of SA's VRNID.

17. Third, SA removed 724,974 records from DL prior to his matching sweeps on account of duplicate SSNs (category E in my ¶ 5). These removals also are problematic. They do not affect SA's sweep #1 – on SSN – but they do affect his sweep #2 and sweep #3. The reason is subtle but clear. The following is an illustrative hypothetical example. Suppose that I am registered to vote as Tom Sager with SSN=3333 and correct birth date.[6] Suppose further that I have two DL entries – one as Tom Sager with SSN=111-22-3333 and the other as Thomas Sager with SSN=111-22-3333, and both have my correct birth date. SA's procedure would delete one of my two DL entries before matching. Whichever one he deletes, he will not match me on his sweep #1 and will need to proceed to sweeps #2 and #3. If he deletes the Tom Sager DL entry, then he will fail to match me on sweeps #2 (DOB+FN+LN) and #3 (DOB+FN+LN+MN) and will put me into VRNID. Only if he deletes the Thomas Sager DL entry will he successfully match me and keep me out of VRNID. In other words, it is not necessary to delete duplicate SSNs from DL/LTC prior to matching sweeps. Retention of SSN duplicates increases the chance of matching voters to true variant names and spellings. Deletion of SSN duplicates tends to inflate the size of VRNID.

[6] 54.6% of the VR database lack full SSN.

18. There remains one SA deletion category (A) referenced in my ¶5 that I have not yet discussed. Category A numbers 287,236 DL entries that SA removed because no license or ID was shown as issued. These removals are unobjectionable on their own but result in no major difference. I inquired whether the State had deleted Category A from the DL database for the May No Match dataset of 588,095 voters. Since it had not, on this account the size of the May No Match database was slightly smaller than it should have been. However, the difference was small – only 8,228 records.

19. SA also made a deletion from the LTC database that will tend to inflate his results. SA deleted 38,919 records because they were recorded as "non-citizens" in the LTC database. Non-citizens are not eligible to vote, and so if these records matched to a VR entry that would either indicate: (a) an ineligible voter that should not be counted as a bona fide voter without ID or (b) a voter who had subsequently naturalized and had an LTC ID and therefore should not be counted as a voter without ID. Either way, SA should not have eliminated these entries. However, the LTC dataset produces so few additional matches that the elimination of this small number of records likely is not nearly as significant as the SA's other deletions of deceased and expired records from the DL dataset.

20. I now discuss the protocols for and results of my attempts to match SA's VRNID database to the DL/LTC databases. In spite of the problems with the construction of SA's VRNID database that I discussed in

the preceding paragraphs, I took SA's VRNID database as he provided it for the purpose of my assignment to match it to the DL/LTC databases. Except that, for this purpose, the VRNID database to match consists of the 1,893,143 "no match" VR cases identified by SA, reduced by exclusion of the 125,015 and 266,151 VR records that SA identified as having "insufficient information" or "ambiguous DL status" (see my ¶ 9).

21. Working with this VRNID database of 1,501,977 voter records, I supervised a re-match to the full DL/LTC databases as they were prior to SA's cleaning.[7] Performing this re-match required first re-appending the SSN data that SA had redacted from the version of VRNID that he delivered to the State.

22. The following four match sweeps were run to match SA's VRNID to both the DL and LTC databases. In each sweep, identical matches on all criteria listed were required.

a. First Name and Last Name and DOB

b. SSN9

c. First Name and Last Name and SSN4

d. SSN4 and DOB

[7] As I indicated above, I remotely supervised OAGLTS personnel in the matching since I was out of the state.

23. In the same manner as described in my June 1, 2012 declaration, I also identified "suspense" entries and over-age-65[8] entries for SA's VRNID database, as I had done for the State's May No Match database.

24. Finally, as a check on SA's cleaning of the DL database, I directed a separate match of VRNID to only those DL records that SA cleaned from the DL database on account of no ID or out-of-state residence, deceased drivers, or expired licenses. The purpose of this separate and more restricted sweep is to discover the matches that SA failed to find as a result of his exclusion of these categories of DL entries. For this separate sweep, all four of the matching criteria at my ¶ 22 were employed. I first discuss the results of this check on the effects of SA's cleaning of the DL database.

25. First, only 746 of VRNID match a no-ID DL entry.

26. Second, SA's removal of DL records for deceased drivers but not for deceased voters makes a more substantial difference. 57,718 of VRNID match DL entries for deceased drivers.

27. Third, 468,775 records in VRNID match a DL record with an expired license.

28. The following table shows a summary of the results of matching all of VRNID to the records that SA cleaned from DL by cleaning category matched.

[8] Over age 65 as of November 6, 2012.

	DPS_EXPIRED	DPS_OTHER_JURIS	DPS_DECEASED	DPS_NoId	Count
	1	1			102,951
	1		1		28,693
	1				337,131
		1			10,320
			1		29,025
				1	746
Not in a "cleaning" category					**993,111**
Total					**1,501,977**

29. The following table expands the preceding table by adding a breakdown by over-age-65 status and by "suspense" status. I understand that voters over the age of 65 are not required to present a photo id in order to vote under SB 14 because they can vote by mail. I further understand that a voter registration in "suspense" means that mail sent to the voter has been returned to sender and that the voter has not yet verified a new address, meaning they may have moved out of state or otherwise become ineligible. Although one can tally from this table 468,775 records in VRNID that match a DL record with an expired license, 110,073 of these are for voters over the age of 65, and 138,426 of these entries are in suspense status.

	65	Suspense	DPS_EXPIRED	DPS_OTHER_JURIS	DPS_DECEASED	DPS_NoId	Count
	1	1	1	1			6,318
	1	1	1		1		5,999
	1	1	1				17,320
	1	1		1			196
	1	1			1		5,095
	1	1				1	4
	1	1					25,792
	1		1	1			5,019
	1		1		1		12,733
	1		1				62,684
	1			1			758
	1				1		14,048
	1					1	36
	1						174,415
		1	1	1			41,492
		1	1		1		3,064
		1	1				64,233
		1		1			2,430
		1			1		2,873
		1				1	156
		1					161,127
			1	1			50,122
			1		1		6,897
			1				192,894
				1			6,936
					1		7,009
						1	550
Not in a "cleaning" category or suspense or over 65							**631,777**
Total							**1,501,977**

30. I turn now to the results of matching VRNID to DL/LTC using the four matching sweeps outlined in my ¶ 22. Exhibit A to this declaration shows the results of this re-match analysis, with separate break-outs for each of the four sweep criteria applied, as well as the overlaps among the four

sweep criteria, and a further breakdown for over age 65 status and suspense status. A partial summary of these results is shown in the table below.

Criteria	Number of VRNID entries
Matched to DPS DL or LTC using any criteria from ¶ 22 (a) – (d)	814,903
Over 65	330,377
Suspense	335,939
Reported ID Number to SOS but did not match any of above	261,887

31. By summing all rows in Exhibit A, one can readily calculate that out of 1,501,977 entries in SA's no-match VRNID database, a total of 1,072,366 voter entries either match one or more of the four matching sweeps in my ¶ 22 (a) – (d), or are over 65 or are in suspense. Furthermore, by summing appropriate rows in Exhibit A, one can calculate that 814,903 of these VRNID entries qualify as matches by application of one or more of my four matching sweeps. This count includes 210,601 matches on full SSN9. By summing appropriate remaining rows, one computes that an additional 140,666 voter entries are over 65, and the balance of 116,797 voter entries are in suspense. That leaves a remainder of 429,611 (= 1,501,977 – 1,072,366) as yet unaccounted for. This number includes 746 VRNID entries that match to the no-ID category in the DL database. An additional 261,887 are voter registrants who reported having an official ID number to the Texas Secretary of State. This leaves only 167,724 of VRNID unaccounted for. Many of these entries (and the larger set of approximately 700,000 entries that do not formally "match" into the DPS database using the criteria of ¶ 22

(a)-(d)) likely do have undetected matches in the DPS database. Over 50% of the VR database lacks full SSN data, but the discussion below and in Exhibit B show that there is a very high rate of name and DOB mis-matches for SSN matching records through alternate names, marriage, or data entry errors.

32. As noted in the table above (my ¶ 30), 261,887 VR entries have an official ID number. I understand that when a voter registers, the Texas Secretary of State ("TXSOS") records the voter's assertion that he/she has state identification. I understand that when in the creation of the "No Match" sets that I discussed in my first declaration, any voter registration entry that includes state ID as belonging to someone was removed from the No Match set. Thus, TXSOS generally took people at their words that they had driver licenses or state IDs when they registered. These individuals (those who did not otherwise match) are listed in the last row of my matching table (my ¶ 30), above, and represent a significant number.

33. As I also noted, 167,724 entries in SA's VRNID database remain unaccounted for by the above procedures. Of these remaining 167,724 voter entries, 31.6% have Spanish surnames, per the appropriate VR data field. This rate is 9.3% higher than the overall Spanish Surname registration rate of 22.25% reported by SA (SA ¶ 33). The difference represents less than 16,000 (9.3% x 167,724) of the unaccounted for voters in VRNID. 16,000 represents approximately 0.55% of the total Spanish surname population of registered voters. On general statistical and data processing grounds, there

are eminently reasonable bases for doubting the meaningfulness of such a relatively small difference. Although such a difference could be "statistically significant" by the rote application of a standard statistical test, neither the statistical grounds for use of such a test nor the justification for interpreting its meaning as implying lesser Hispanic access to State identification has been established to a reasonable certainty, in my opinion.

34. For example, if Hispanics have even a slightly higher rate than non-Hispanics for holding the appropriate federally issued identification, the ethnic difference could be mooted. Also, uncertainties about the quality of data (about which I have much more to say below) could turn 16,000 into a rounding error. If Hispanics have even a slightly higher probability than non-Hispanics of name variants or misspellings, SA's match sweeps will tend to place relatively more of them into his unmatched VRNID database.

35. The operation of such a selection bias can be demonstrated for females in SA's VRNID database. Females appear to be over-represented in VRNID. SA's sweep matching criteria are less likely to match females than males. One reason is that many females change their last names when they marry and therefore are more likely to have different last names in VR than in DL/LTC. SA's matching criteria that require identical first and last names will assign such females incorrectly to VRNID. The effect can be substantial. In an empirical analysis of VRNID that I discuss in greater detail in Exhibit C. I took a random sample of 1,000 entries from a subset of VRNID known to

have matches in DL/LTC with very high confidence. 650 of the 1,000 are female and only 350 are male. Such a disparity cannot be explained by chance. I counted approximately 207 females who appear to have changed their last names. Further analysis in Exhibit C confirms that in matching criteria that do not require first and last name matching, there is a selection bias for male matches.

36. As a statistician, I often perform statistical tests of the difference between two percentages. Those tests are designed to determine whether a difference is real and not attributable to chance; the tests are not designed to determine whether the difference is meaningful¸ nor to determine the cause of the difference. If one has enough data, nearly every difference will test as real. Meaningfulness must be judged by other criteria. Moreover, the validity of such tests depends upon satisfying the assumptions upon which they depend. Furthermore, if one is not careful, then the cause of a statistically significant difference may not be what one assumes. For example, a higher rate of Spanish than non-Spanish surnames among unmatched voters may not result from less access to state identification by those with Spanish surnames but may result from a higher rate of data errors in Spanish surnames. Above, I cite the over-representation of females in VRNID relative to males. Does this imply that Texas females have less access to State identification than males? No, it just means that there are more "data errors" for females in the voter and driver license databases on

account of many Texas females following the social custom of adopting their husbands' last names in marriage.

37. In the current case, both SA and I have reported most of our tallies with a precision that belies potentially large uncertainties about the quality of the data that we used and limitations in the available means for measurement. I think this point about data uncertainties is quite important. In fact, I have conducted an investigation into the quality of some of the data in VRNID. I will briefly mention five data problems here, but leave the detailed discussion to Exhibit B.

38. Name variants. Many names have variants and nicknames. "Thomas", "Tom", "Tommy" are all variants of the same name. The requirement of an identical match on name will fail to match if the VR database has a different variant from the DL/LTC database.

39. Name misspellings. A clerical error in reading or typing a name can result in a misspelling and consequent mismatch. The requirement of an identical match on name will fail to match me if I am in the VR as "Tomas" and in DL as "Thomas".

40. Married female name changes. If my wife had a driver's license before she married me and registered to vote after marriage and changed her last name after marriage, then she would be in DL as "Alexander" and in VR as "Sager". Consequently, the requirement of an identical match on name will fail to match her.

41. Date of Birth. Clerical errors in reading and/or data input may result in dates of birth that are slightly different between VR and DL/LTC. For example, a hand-written date of "4/19/46" might be misread as "4/14/46" if the "9" is misread as a "4".

42. Differential name error rates by ethnicity. To a lay mind, it is possible to think of reasons why Spanish surnames may be more prone to data entry error than non-Spanish surnames. For example, a data entry clerk may not be familiar with Spanish surnames and perhaps be more likely to misspell or misinterpret them. For example, is a hand-written "DE LA CRUZ" one word, two words, or three words – or should it be "DE LA CRUS?"

43. Exhibit B discusses a quantitative analysis of data issues in SA's no-match VRNID by using a random sample of 1000 entries taken from a subpopulation of VRNID entries that are known with high confidence to have been correctly matched to the same people in DL/LTC. Since this sample of 1000 people have been correctly matched, their first names, last names, and dates of birth should be the same in VR and in DL/LTC. Yet many are not. 340 fail to match on exact first name, 388 fail to match on exact last name, and 368 fail to match on exact date of birth. To be sure, such high error rates should not be extrapolated to the VR and DL/LTC databases generally, since the VRNID database was formed from VR entries whom SA had difficulty matching on these criteria. However, this sample of 1,000 known true matches provides a laboratory for testing the prevalence of

VRNID data issues without the otherwise legitimate concern that differences in names and DOBs might be true differences resulting from the mismatch of different people. These are the same people, yet their VR data is often not the same as their DL/LTC data. The effect of data errors in first name, last name, and/or date of birth is to reduce the chance of finding true matches and hence to inflate the size of VRNID.

44. Exhibit D discusses the expected matching error rates on the matching criteria other than SSN9.

I swear the foregoing is true and correct to the best of my knowledge.

Dated: July 10, 2012

Respectfully submitted.

Thomas W. Sager

Exhibit A

	65	SUSPENSE	FnLnDob	LTC	SSN9_Combo	SSN4_Name	SSN4_DOB	VR Entry Count
Total								**1,501,977**
SSN4 DOB \							1	79,529
SSN4 Name \						1		8,177
SSN4 Name \SSN4 DOB \						1	1	1,319
SSN9 \					1			729
SSN9 \SSN4 DOB \					1		1	10,088
SSN9 \SSN4 Name \					1	1		2,074
SSN9 \SSN4 Name \SSN4 DOB \					1	1	1	185
LTC \				1				831
LTC \SSN4 DOB \				1			1	114
LTC \SSN4 Name \				1		1		2
LTC \SSN4 Name \SSN4 DOB \				1		1	1	1
LTC \SSN9 \SSN4 DOB \				1	1		1	9
LTC \SSN9 \SSN4 Name \				1	1	1		1
LTC \SSN9 \SSN4 Name \SSN4 DOB \				1	1	1	1	1
FnLnDob \			1					212,714
FnLnDob \SSN4 DOB \			1				1	1,810
FnLnDob \SSN4 Name \			1			1		246
FnLnDob \SSN4 Name \SSN4 DOB \			1			1	1	56,369
FnLnDob \SSN9 \			1		1			17
FnLnDob \SSN9 \SSN4 DOB \			1		1		1	97
FnLnDob \SSN9 \SSN4 Name \			1		1	1		22
FnLnDob \SSN9 \SSN4 Name \SSN4 DOB \			1		1	1	1	89,466
FnLnDob \LTC \			1	1				1,932
FnLnDob \LTC \SSN4 DOB \			1	1			1	14
FnLnDob \LTC \SSN4 Name \			1	1		1		1
FnLnDob \LTC \SSN4 Name \SSN4 DOB \			1	1		1	1	325
FnLnDob \LTC \SSN9 \SSN4 DOB \			1	1	1		1	1
FnLnDob \LTC \SSN9 \SSN4 Name \			1	1	1	1		1
FnLnDob \LTC \SSN9 \SSN4 Name \SSN4 DOB \			1	1	1	1	1	695
Suspense \		1						116,797
Suspense \SSN4 DOB \		1					1	17,095
Suspense \SSN4 Name \		1				1		1,679
Suspense \SSN4 Name \SSN4 DOB \		1				1	1	235
Suspense \SSN9 \		1			1			170
Suspense \SSN9 \SSN4 DOB \		1			1		1	2,830
Suspense \SSN9 \SSN4 Name \		1			1	1		547
Suspense \SSN9 \SSN4 Name \SSN4 DOB \		1			1	1	1	61
Suspense \LTC \		1		1				297
Suspense \LTC \SSN4 DOB \		1		1			1	26
Suspense \LTC \SSN4 Name \		1		1		1		2
Suspense \LTC \SSN9 \SSN4 DOB \		1		1	1		1	5
Suspense \FnLnDob \		1	1					90,367
Suspense \FnLnDob \SSN4 DOB \		1	1				1	475
Suspense \FnLnDob \SSN4 Name \		1	1			1		54
Suspense \FnLnDob \SSN4 Name \SSN4 DOB \		1	1			1	1	15,637
Suspense \FnLnDob \SSN9 \		1	1		1			5
Suspense \FnLnDob \SSN9 \SSN4 DOB \		1	1		1		1	28
Suspense \FnLnDob \SSN9 \SSN4 Name \		1	1		1	1		5
Suspense \FnLnDob \SSN9 \SSN4 Name \SSN4 DOB \		1	1		1	1	1	27,554
Suspense \FnLnDob \LTC \		1	1	1				890
Suspense \FnLnDob \LTC \SSN4 DOB \		1	1	1			1	4
Suspense \FnLnDob \LTC \SSN4 Name \SSN4 DOB \		1	1	1		1	1	123
Suspense \FnLnDob \LTC \SSN9 \SSN4 DOB \		1	1	1	1		1	1
Suspense \FnLnDob \LTC \SSN9 \SSN4 Name \SSN4 DOB \		1	1	1	1	1	1	332
65 \	1							120,810
65 \SSN4 DOB \	1						1	15,019
65 \SSN4 Name \	1					1		8,818
65 \SSN4 Name \SSN4 DOB \	1					1	1	456
65 \SSN9 \	1				1			1,089
65 \SSN9 \SSN4 DOB \	1				1		1	5,739
65 \SSN9 \SSN4 Name \	1				1	1		5,098
65 \SSN9 \SSN4 Name \SSN4 DOB \	1				1	1	1	126
65 \LTC \	1			1				433
65 \LTC \SSN4 DOB \	1			1			1	41
65 \LTC \SSN4 Name \	1			1		1		14
65 \LTC \SSN4 Name \SSN4 DOB \	1			1		1	1	1
65 \LTC \SSN9 \	1			1	1			2
65 \LTC \SSN9 \SSN4 DOB \	1			1	1		1	13
65 \LTC \SSN9 \SSN4 Name \	1			1	1	1		8
65 \FnLnDob \	1		1					50,696
65 \FnLnDob \SSN4 DOB \	1		1				1	174
65 \FnLnDob \SSN4 Name \	1		1			1		71
65 \FnLnDob \SSN4 Name \SSN4 DOB \	1		1			1	1	11,730
65 \FnLnDob \SSN9 \	1		1		1			26
65 \FnLnDob \SSN9 \SSN4 DOB \	1		1		1		1	12
65 \FnLnDob \SSN9 \SSN4 Name \	1		1		1	1		12
65 \FnLnDob \SSN9 \SSN4 Name \SSN4 DOB \	1		1		1	1	1	47,196
65 \FnLnDob \LTC \	1		1	1				1,128
65 \FnLnDob \LTC \SSN4 DOB \	1		1	1			1	4

65 \FnLnDob \LTC \SSN4 Name \SSN4 DOB \	1		1	1		1	1	159
65 \FnLnDob \LTC \SSN9 \	1		1	1	1			2
65 \FnLnDob \LTC \SSN9 \SSN4 Name \SSN4 DOB \	1		1	1	1	1	1	780
65 \Suspense \	1	1						19,856
65 \Suspense \SSN4 DOB \	1	1					1	1,312
65 \Suspense \SSN4 Name \	1	1				1		859
65 \Suspense \SSN4 Name \SSN4 DOB \	1	1				1	1	37
65 \Suspense \SSN9 \	1	1			1			302
65 \Suspense \SSN9 \SSN4 DOB \	1	1			1		1	1,248
65 \Suspense \SSN9 \SSN4 Name \	1	1			1	1		1,238
65 \Suspense \SSN9 \SSN4 Name \SSN4 DOB \	1	1			1	1	1	31
65 \Suspense \LTC \	1	1		1				83
65 \Suspense \LTC \SSN4 DOB \	1	1		1			1	4
65 \Suspense \LTC \SSN4 Name \	1	1		1		1		1
65 \Suspense \LTC \SSN9 \SSN4 DOB \	1	1		1	1		1	8
65 \Suspense \LTC \SSN9 \SSN4 Name \	1	1		1	1	1		4
65 \Suspense \FnLnDob \	1	1	1					19,574
65 \Suspense \FnLnDob \SSN4 DOB \	1	1	1				1	27
65 \Suspense \FnLnDob \SSN4 Name \	1	1	1			1		16
65 \Suspense \FnLnDob \SSN4 Name \SSN4 DOB \	1	1	1			1	1	2,894
65 \Suspense \FnLnDob \SSN9 \	1	1	1		1			13
65 \Suspense \FnLnDob \SSN9 \SSN4 DOB \	1	1	1		1		1	7
65 \Suspense \FnLnDob \SSN9 \SSN4 Name \	1	1	1		1	1		5
65 \Suspense \FnLnDob \SSN9 \SSN4 Name \SSN4 DOB \	1	1	1		1	1	1	12,450
65 \Suspense \FnLnDob \LTC \	1	1	1	1				420
65 \Suspense \FnLnDob \LTC \SSN4 DOB \	1	1	1	1			1	1
65 \Suspense \FnLnDob \LTC \SSN4 Name \	1	1	1	1		1		1
65 \Suspense \FnLnDob \LTC \SSN4 Name \SSN4 DOB \	1	1	1	1		1	1	61
65 \Suspense \FnLnDob \LTC \SSN9 \	1	1	1	1	1			1
65 \Suspense \FnLnDob \LTC \SSN9 \SSN4 Name \SSN4 DOB \	1	1	1	1	1	1	1	267
Remaining Entries								**429,611**
Registrants who reported Official ID Number to SOS								261,887
Remainder								**167,724**

Exhibit B

Analysis of a Random Sample of 1,000 VRNID Entries
that Match on SSN9 to the DPS Dataset

At various places in the main body of my second declaration, I refer to further empirical analysis of SA's VRNID database. I discuss those empirical analyses here in this Exhibit. At ¶ 22, I list the four matching criteria that I used to attempt to match SA's VRNID database to the DL/LTC databases. At ¶ 30, I report that I was able to match 814,903 entries in SA's VRNID on one or more of the four criteria. In order to check the accuracy of those matches and to learn more about the VRNID entries matched by those four criteria, I directed that a random sample of entries be selected for each one of the four matching criteria. In each case the random sample would consist of 1,000 entries drawn at random from the set of VRNID entries that had been matched to DL/LTC by using one of the criteria. Since each sample is large and drawn at random, it should be representative of the subpopulation of VRNID entries that were matched by its criterion. Furthermore, each sample is just small enough to permit some (laborious) manual examination.

I now discuss each of the four samples. I discuss the SSN9Combo sample at length because of its special significance.

SSN9Combo Sample

The SSN9Combo Sample is a random sample of 1,000 entries from a subset of SA's VRNID database. OAGLTS drew this sample at my direction from the entries in SA's VRNID that OAGLTS had matched to the DPS drivers license (DL) and license-to-carry (LTC) databases on the basis of identical SSN9. I had this sample drawn with the intention of manually examining all 1,000 records in the sample. I want to develop an understanding of problems that may arise in matching records from different databases for the same entries. These problems may suggest reasons that might account in part for the much larger size of SA's no-match list (VRNID, with 1,501,977 ultimate entries) compared with the State's May No Match list (with 588,095 entries). Ideally, information for the same voter should be the same in different databases. It often is not. Since the 1,000 records in SSN9Combo are a random sample, the problems that I uncover in my analysis of them should be representative of problems with the 31,606 records from which the sample was drawn.

Since an SSN9 should be unique to one individual, the presumption is that all 31,606 individuals were matched to themselves in the DL and LTC databases, absent recording errors, confusion, or fraud. In particular, the presumption is that the 1,000 entries in the SSN9Combo sample were matched to themselves in the DL and LTC databases. It is therefore instructive to compare *other* data that these entries have in the voter registration (VR) database with their corresponding *other* data in the DL and LTC databases. Do these presumably identical people have the same names, birthdates, sex, and residence information in DL and LTC databases as in the VR database? Answers to such questions can point to the scope of the difficulties in using simple computer matching, like SA and the OAGLTS and I used, to pair VR records with DL and LTC records.

The matching criteria that SA, OAGLTS, and I used require identical matches. Identical matching for SSN9 is clearly proper, given that SSN9s should be unique. However, when one matches names, misspellings may create the appearance of mismatches that are not real mismatches. There are other issues that contribute to inflated mismatch counts, as well. Examination of the 1,000 SSN9Combo Sample will help clarify these issues.

First, I compared the 1,000 entries on the basis of sex. VR agreed with DL/LTC for 959 of the 1000 entries in SSN9Combo. Of the 41 mismatches, the VR data had 35 records with missing or unspecified sex. The DL/LTC data were complete, with a valid M or F for all 1,000 entries. Of the 6 remaining M-F or F-M mismatches, 2 have the same names and residences, 2 are perhaps two married couples using the same SSN9, and 2 remain problematic.

Then I matched the 1,000 entries in SSN9Combo Sample on the basis of identical date-of-birth, last name, and first name. The results for all 8 combinations are shown below.

Matching 1,000 SSN9Combo Sample to DL/LTC (“1” indicates exact match; “0” indicates mismatch)

LAST	FIRST	DOB	Count
1	1	1	0
1	1	0	306
1	0	1	317
1	0	0	37
0	1	1	281
0	1	0	25
0	0	1	34
0	0	0	0
		Total =	1000

Two initial observations: First, none of the 1,000 entries fails to match on all three criteria simultaneously. Second, none of the 1,000 entries has identical matches on all three criteria (DOB, last and first names). SA placed entries into his VRNID no-match dataset because he found no matches on one or more criteria. So it should not be inferred that the mismatch rate on DOB, last and first names shown here will be representative of the entire VR data generally. The mismatch rates for the VR data in general should be much lower than the rates found in SA’s VRNID data. However, the mismatch rates found in my sample should be representative of the 31,606 in SA’s VRNID data that OAGLTS could match to DL/LTC data by SSN9, in spite of SA’s claim that none of his VRNID records match on SSN9.

A total of 340 entries fail to match on last name. I manually inspected each of these 340 entries and judged 125 of the last name mismatches to be variant spellings or misspellings. I further judged 207 of the remainder to be likely instances of married women having adopted their husbands’ last names. All 207 are females who have either identical first names or identical dates-of-birth. That leaves 8 last name mismatches, of whom all but one have identical birth dates. The one remaining entry has the same first name and a birth date that differs by one day. I judge all 340 to be likely true matches.

A total of 660 entries matched exactly on last name, but failed to match exactly on at least one of first name and/or birth date. 317 of these matched exactly on birth date but not on first name. I manually examined all 317 and found 216 first names to be likely variant spellings or misspellings, 51 to have interchanged first and middle names,[1] 39 with the same residence, and 11 unmatched by these secondary means. On the basis of exact match on birth date and first name and/or residence corroboration on most of them, I judge these 317 to be likely true matches.

306 entries matched exactly on last name and on first name, but not on birth date. 92 of these have suspect birth dates in the VR file.[2] 80 of the 92 matched on residence. The remaining 214 birth date discrepancies are too numerous to characterize confidently and tabulate by category. They include differences of 1 in the birth year or birth day, or transpositions of birth month with birth day, and a number of possible confusions of one digit for another.[3] Most birth date discrepancies appear to be of this type. I manually matched 148 of the 214 on residence (street name).

Summary observations: Almost all of the 1,000 SSN9Combo Sample can be corroborated as true matches by matching on alternative criteria than SSN9. For example, there are 306 exact matches on first and last names; there are an additional 632 exact matches on birth date. 25 of the 62 remainder do match on first names and of these 9 match on variant spelling of last name; 15 are females who likely changed last names when they married; and 1 has birth dates that differ by one day. Of the final 37, all do match on last name and of these, 32 also match on street name or variant spelling of first name or FN/MN swap; 5 are unexplained, although 2 of the 5 have identical birth months and birth days and 1 has a missing birth date.

Thus, we can be reasonably confident that almost all of the 1000 entries in the SSN9Combo Sample really are validly matched to themselves in the DL/LTC databases. In spite of this, there are many problems with matching on criteria other than SSN9.

- 34% of last names of the same entries fail to match exactly. About 37% (=125/340) of last name mismatches are attributable to probable misspellings or variant spellings. A further 61% (=207/340) of last name mismatches are likely attributable to females who change their last name when they marry.

- 39% of first names of the same entries fail to match exactly. About 65% (=254/388) of first name mismatches are attributable to probable misspellings or variant spellings. A further 15% (=59/388) of first name mismatches are likely

[1] If a person often goes by middle name, s/he may provide the actual first name for one database and the middle name for the other.
[2] 20 have birth dates of 1-1-1900, 4 with 1-1-1901, and 68 with 1-1-1911.
[3] For example, misreading handwritten “2” as “7”, “7” as “9”, “5” as “9”, “4” as “9”, a “/” written to divide parts of a date as a “1”, etc.

attributable to swapping first names with middle names.

- 37% of birth dates of the same entries fail to match exactly. I did not attempt to tally these mismatches into categories, but most may be simple clerical errors.

- The VR records contain substantial suspect birth date data – 9% (=92/1000) of the SSN9Combo Sample have birth dates of 1-1-1900, 1-1-1901, or 1-1-1911.

- The VR records also contain numerous missing sex information – 4% (=41/1000) of the SSN9Combo Sample.

- The preceding two bullets suggest that the VR data may be less complete and/or less accurate in some respects than the DL/LTC data. As a test of this hypothesis, I examined the 125 records in the SSN9Combo Sample for which I had previously determined that misspellings and/or variant spellings accounted for the failure of a last name match. I determined that 20 of the 125 records permitted attribution of responsibility of last name mismatch to either the VR data or to the DL/LTC data with reasonable confidence – in my judgment. Of the 20, I attributed responsibility to VR for 12 records and to DL/LTC for 8 records.[4] Although this test is suggestive, the sample size is too small to draw definitive conclusions.

- As I was checking the entries for misspellings, I observed misspellings more frequently among Spanish surnames than among non-Spanish surnames. I wished to examine this possibility further. There are 350 males in the SSN9Combo sample. I manually examined each male surname and classified it as Spanish, "difficult foreign", or "other" – obtaining counts of 97, 8, and 245, respectively. I excluded females in order to avoid the complications of married name changes. I then tallied the number of each group that I had previously classified as not matching on last name because of misspelling, obtaining counts of 23, 2, and 8, respectively. Thus, I classified 24% (= 23/97) of male Spanish surnames to be misspelled, 25% (= 2/8) of male difficult foreign names, and 11% (= 28/245) of male "other" surnames. The difference between Spanish and "other" misspelling rates is easily big enough to pass the standard test for statistical significance if my determinations are accepted.

Finally, I note the male/female representation in this sample for use in Exhibit C: 650 females, 350 males.

SSN4 + DOB Sample

The SSN4+DOB Sample is a random sample of 1,000 entries from a subset of SA's VRNID database. OAGLTS drew this sample at my direction from the entries in SA's

[4] For example, a last name was "WHITACFE" in VR and "WHITACRE" in DL/LTC (assigned responsibility to VR). Another example is "CLAYTON" in VR and "CLAYTOR" in DL/LTC (assigned responsibility to DL/LTC).

VRNID that OAGLTS had matched to the DPS drivers license (DL) and license-to-carry (LTC) databases on the basis of identical SSN4 and DOB.

In the SSN4+DOB sample, of course, all 1,000 entries have the same birth date in the VR database as in the DL/LTC database. However, 640 entries have different last names, and 550 have different first names. After manual examination of the 1000 records, I determined that 224 of the 640 last name mismatches were likely due to last name misspellings or variants, and a further 330 likely due to female last name changes upon marriage. I also determined likely first name misspellings and variants (336) and interchanges of first name with middle name (24). In view of the fact that these 1,000 entries have already been matched on the basis of identical SSN4 and DOB, if we require one more field match – on last name and/or first name – in order to confirm a record match, then 928 of the 1000 records in SSN4+DOB are confirmed matches. This yields a potential false match rate for the SSN4+DOB criterion of about 7% (= 72 / 1000). This error rate is within the range of 5% - 10% arrived at later in Exhibit D by theoretical and other empirical considerations.

Finally, I note the male/female representation in this sample for use in Exhibit C: 673 females, 327 males.

SSN4 + FN + LN Sample

The SSN4+FN+LN Sample is a random sample of 1,000 entries from a subset of SA's VRNID database. OAGLTS drew this sample at my direction from the entries in SA's VRNID that OAGLTS had matched to the DPS drivers license (DL) and license-to-carry (LTC) databases on the basis of identical SSN4 and first name and last name.

In the SSN4+FN+LN sample, of course, all 1,000 entries have the same first and last names in the VR database as in the DL/LTC database. However, all 1,000 entries have different dates of birth (otherwise they would have matched in SSN4 + DOB).

An inspection suggests that at least a large majority of this sample are probably correctly matched: Most of the DOB discrepancies appear to be of the type discussed above for the SSN9Combo Sample; most of the street names appear to match; and they are already known to have the same SSN4, first and last names.

Finally, I note the male/female representation in this sample for use in Exhibit C: 513 females, 487 males.

FN + LN + DOB Sample

The FN+LN+DOB Sample is a random sample of 1,000 entries from a subset of SA's VRNID database. OAGLTS drew this sample at my direction from the entries in SA's VRNID that OAGLTS had matched to the DPS drivers license (DL) and license-to-carry (LTC) databases on the basis of identical first name and last name and date of birth.

In the FN+LN+DOB sample, of course, all 1,000 entries have the same first and last names and birth dates in the VR database as in the DL/LTC database.

Checking Exhibit A shows that about 60% of the sample match on SSN9 or SSN4. Cursory inspection of street names shows that most are the same; about 80% match on city and about 75% match on zip code.

Finally, I note the male/female representation in this sample for use in Exhibit C: 497 females, 503 males.

Exhibit C

An Example of Selection Bias in SA's VRNID Database

In the main body of my second declaration at my ¶ 35, I briefly referred to the over-representation of females in SA's VRNID. Females appear to be over-represented relative to males in VRNID because SA's matching criteria for his three sweeps are selectively biased against matching females. Consequently, relatively more females than males are assigned to VRNID. At least one reason for the over-representation of females in VRNID is not that females have less access to State identification than males, but that many of them change their last names when they marry. Thus, females are more likely than males to have last names that differ between the voter registration rolls and the driver license database. Any matching criterion that requires identical last names in VR and DL/LTC will therefore be less likely to find matches for females than for males.

Is there empirical evidence for the assertions of the preceding paragraph? Yes. I had four random samples selected from SA's VRNID, with each sample drawn from a different subset of VRNID that was matched to DL/LTC by a different matching criterion. The SSN9Combo Sample was matched to DL/LTC on the basis of identical SSN9s – but did not use names. There are 650 females in SSN9Combo and only 350 males, rather than a more nearly even split that one might expect. The disparity is significant and impossible to attribute reasonably to chance. Why such a big difference? I counted 207 plausible instances of females in SSN9Combo who had different last names in VR than in DL/LTC – differences in last names that were not likely simple misspellings or name variants. There are only 8 males with different last names that are not simple misspellings or name variants.[1]

The situation is much the same in the SSN4+DOB Sample. This sample was matched to DL/LTC on the basis of identical SSN4 and DOB – but did not use names. There are 673 females and only 327 males. Again, the disparity is huge and impossible to attribute reasonably to chance. Again, I counted many female last name changes – 330.

On the other hand the situation is much different in the SSN4+FN+LN Sample and in the FN+LN+DOB Sample. The sex split is more nearly even in these two samples – respectively, 513 F / 487 M and 497 F / 503 M. Why is the sex split nearly even in these two samples and so disparate in the other two? Because the matching criteria in these two samples uses exact first name and last name. Therefore, these matching criteria exclude females who have changed their last names between the time of attaining a driver license and the time of registering to vote (in whichever order).

This discussion of selection bias against matching females is presented as a cautionary tale. The implication is that a selection bias against matching Hispanics might also exist if Hispanics are more likely than non-Hispanics to have errors in recording their first and/or last names. In my discussion of SSN9 matching (my Exhibit B), I presented some suggestive empirical evidence in favor of such a bias in VRNID. One can think of possible reasons for the existence of more Hispanic name errors. For example, a data

---

[1] Of these 8, 3 had interchanges of first and last name or middle and last name, the others all have the same birth dates and first names.

entry clerk may be more likely to misspell or misinterpret them. As another example, when a Hispanic person is asked to provide a surname, he may misunderstand *which* of his surnames he should give.

Exhibit D

False Match Error Rates on the Matching Criteria

In ¶ 22 (a) – (d) of my second declaration, I set forth four matching criteria for my attempt to find additional matches in SA's VRNID. As discussed in my Exhibit B, there are good reasons to believe that when matching on SSN9, the error rate of false matches is close to zero. In this section, I discuss the error rate of false matches for the other three matching criteria.

SSN4 + DOB

Although SSN9 is theoretically unique, neither SSN4 nor DOB is unique. The combination SSN4+DOB is not even unique. A few approximate theoretical computations may illuminate the extent of problems with matching on SSN4 and DOB.

Let us suppose a DL database of approximately 20,000,000 distinct drivers. There are 10,000 distinct possible combinations of the 4-digits in SSN4. If we assume that SSN4s are distributed approximately evenly throughout the population, then approximately 2,000 distinct drivers (= 20,000,000 / 10,000) have the same SSN4 as I do (or any particular person does).

However, the distribution of birth date is not even throughout the population, as there are fewer old people than young. However, let us assume as a first approximation that most drivers are between 20 and 70 years of age and that birth dates are roughly evenly distributed throughout that range. Under this assumption, there are approximately 365 x 50 = 18,250 roughly evenly distributed birth dates. For simplicity, I will round this result to 20,000. So I might expect to find roughly 1,000 distinct drivers (= 20,000,000 / 20,000) with the same birth date as mine.[1]

How many distinct drivers may I expect to have both the same SSN4 and birth date as mine? If SSN4 and birth dates are distributed independently of each other, then there are roughly 10,000 x 20,000 = 200,000,000 evenly distributed unique possible combinations of SSN4+DOB. Therefore, I may expect to have roughly a 10% chance (= 20,000,000 / 200,000,000) of finding one or more other drivers with the same combination of SSN4+DOB as mine. Even with duplicates, I may still be properly matched. For example, if I am one of two entries with the same SSN4+DOB and the matching mechanism chooses between us with equal probability, then I have a 50% chance of being the one chosen. Thus, I may have less than 10% chance of being unmatched – perhaps as low as 5%. That is, I may expect roughly 5%-10% of matches on SSN4+DOB to be false matches and 90%-95% to be true matches.

[1] In fact, there are more than 100,000 DL entries for every year from 1930 through 1996, with counts exceeding 10,000 going back to 1913, and with each of the years 1951-1993 having between 300,000 and 500,000 entries. The effect of having uneven counts throughout the years is to increase the chance of false matches; the effect of including more than 50 years of birth dates is to decrease the chance of false matches.

In fact, the preceding rough calculations seem to be confirmed empirically. Within the DL database, approximately 11.5% of the entries have duplicates with the same SSN4+DOB – without attempting to exclude duplicates for the same people.[2] Moreover, about 90% of these duplicates in fact appear only twice in the DL database. Thus, the correct match may still be selected with a probability of about 94% (about 100% – 11.5÷2%) Thus, it may be reasonable to expect that approximately 6% of VR entries that are matched by SSN4+DOB will be false matches.

SSN4 + FN + LN

It is more difficult to calculate reasonable theoretical values for the chance of a false match on SSN4+FN+LN. There are once again 10,000 distinct possible SSN4s. However, both the number and the frequency of first and last names are not known to me. As with the SSN4+DOB criteria, we may perhaps obtain reasonable empirical estimates of the false match rate by consulting the DL database. Within the DL database, approximately 7% of the entries have duplicates with the same SSN4+FN+LN. Moreover, about 97% of these duplicates in fact appear only twice in the DL database. Thus, the correct match may still be selected with a probability of about 96% (about 100% – 7÷2%) Thus, it may be reasonable to expect that approximately 4% of VR entries that are matched by SSN4+FN+LN will be false matches.

FN + LN + DOB

It is still more difficult to calculate reasonable theoretical values for the chance of a false match on FN+LN+DOB. As with the other matching criteria, we may perhaps obtain reasonable empirical estimates of the false match rate by consulting the DL database. The outcome is similar to the other matching criteria. Within the DL database, approximately 7% of the entries have duplicates with the same FN+LN+DOB. Moreover, about 98% of these duplicates in fact appear only twice in the DL database. Thus, the correct match may still be selected with a probability of about 96% (about 100 – 7÷2) Thus, it may be reasonable to expect that approximately 4% of VR entries that are matched by FN+LN+DOB will be false matches.

[2] The effect of data errors in recording DOB and/or SSN4 is to increase the number of false negatives (true matches that are missed by the matching criteria).