Los Angeles County Voter Registration Audit Report, 2020 Election Cycle

R. Michael Alvarez, Silvia Kim and Jian Cao
California Institute of Technology

June 14, 2020

1 Audit Summary: June 14, 2020

We have developed a methodology to audit the Los Angeles County voter registration database for unusual activity. This change detection algorithm looks for record changes, records added, and records dropped. It produces a detailed report which we provide to the Los Angeles County Registrar-Recorder/County Clerk (LA RR/CC) on a regular basis. Between June 14, 2019, and June 8, 2020, the date of the most recent data we received, our algorithm and interquartile range (IQR) analysis have found 36 'events.' These events have been reported to LA RR/CC.

2 Brief Description of Methodology

We have been examining daily changes in the Los Angeles County voter database, starting June 14, 2019. Our methodology first matches the consecutive snapshots of voter registration (VR) data (snapshot $t-1$ and snapshot t), then focuses on analyzing three types of changes: Added Records, Dropped Records, and Changed Records (Figure 1). This methodology is

discussed in detail in Kim, Schneider and Alvarez (2018).[1]

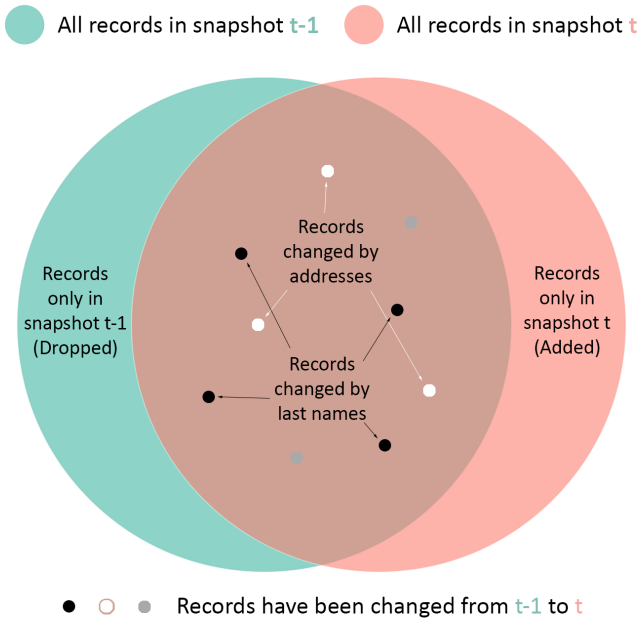

Figure 1: Added, Dropped, and Changed Records

Added Records: VR records that did not exist in snapshot $t-1$ but can be found in snapshot t. These records are represented by the lighter red crescent in the right part of Figure 1.

Dropped Records: VR records from snapshot $t-1$ that can no longer be found in snapshot t. These records correspond to the lighter green crescent in the left part of the figure.

Changed Records: VR records from snapshot t that have been identified (using partial string matching or ID-only matching) as existing records in snapshot $t-1$, but at least one of the fields (e.g., address, last name) changed during the period $t-1$ to t. These changed records are illustrated as black,

[1]Seo-young Silvia Kim, Spencer Schneider, and R. Michael Alvarez, "Evaluating the Quality of Changes in Voter Registration Databases", *American Politics Research*, 2018, https://doi.org/10.1177/1532673X19870512.

white, and grey points in the oval-shaped area located at the center of the figure.

The remaining records that match exactly between snapshot $t-1$ and snapshot t, in all fields, are Exact Matches (usually more than 99% of a snapshot).

We match the snapshots using record linkage techniques. This method enables the linking of two records that are unmatched but point to the same entity. It happens because a person may re-register or update their information after moving, marrying, or similarly making changes to their information.

For the record linkage, we use the following selection of important variables (ADGN variables from Ansolabehere and Hersh, 2017) in the partial string matching.[2]

- Name: name_last, name_first
- Date of Birth: birth_date
- Address: house_number, zip
- Phone Number: phone_1

Matching is performed by `fastLink` (Enamorado, Fifield, and Imai 2019).[3] Default partial string matching parameters are employed. That is, 85% is the lower bound for the posterior probability of a match that will be accepted. Each column has the lower bound for a partial agreement as 88%.

We provide several visual examinations of the results of the linkage process, which include dynamics of dropped records, plots of change rates, directions of voter status changes, directions of party changes, etc. In addition to the visual examinations, we use the interquartile range (IQR) method to detect anomalies. The IQR method aligns the data by size, and then finds points that fall outside of the $[Q1-3\times(Q3-Q1), Q3+3\times(Q3-Q1)]$ range, using the first quartile (Q1) and the third quartile (Q3). While usually, the IQR factor is 1.5, we use 3 to identify events more cautiously—i.e., detect only extreme outliers.

[2]Stephen Ansolabehere and Eitan Hersh, "ADGN: An Algorithm for Record Linkage Using Address, Date of Birth, Gender, and Name", *Statistics and Probability*, 4, 1-10.

[3]Ted Enamorado, Benjamin Fifield, and Kosuke Imai, "Using a Probabilistic Model to Assist Merging of Large-scale Administrative Records", *American Political Science Review*, 113, 2, 353-371.

3 Summary of the Results

Between June 14, 2019, and June 8, 2020, our change detection algorithm and IQR analysis have found 36 events. Our team has provided a detailed report to LA RR/CC regarding these 36 events.

We will continue to update this report in the 2020 election cycle as we receive additional data.