

# Privacy Concerns in New York City Elections

Jesse T. Clark, Lindsey Cormack, Sam Wang

September 20, 2021  
Report EIL-2021-101

A report from the



Princeton University, Green Hall, Washington Road, Princeton, NJ 08542  
[democracy.princeton.edu](https://democracy.princeton.edu)

## Privacy Concerns in New York City Elections

Jesse T. Clark, Lindsey Cormack, and Sam Wang

**Summary:** Two factors to enhance confidence in democracy are the secret ballot and auditable public election records. Normally, it is possible to have both, because votes are aggregated into precincts so that individual votes are anonymized. Here we report that because multi-precinct aggregation was not done in New York City’s primary election of June 22, 2021, it was possible to re-identify hundreds of voters’ ballot choices. Cast-vote records contained hundreds of electoral districts with only one ballot. In voter file lists, we have found that in 378 of these precincts, exactly one person was reported as having voted. In these cases, we can match the person with the vote cast, thus re-identifying their secret ballot. We recommend that this re-identification risk be removed by resuming the practice of aggregating low-turnout precincts. In this way, privacy concerns can be balanced against transparency requirements, thus retaining the ability to monitor and ensure free and fair elections in New York City.

### Introduction

In New York City, data reporting requirements mandate the release of vote records after each election. To comply with this requirement, on August 18, 2021 the city Board of Elections released vote records for the primary election of June 22. The records released from the first ranked-choice voting election are called cast vote records (CVRs), listing the ranked choices each voter cast in the Democratic or Republican primary election as well as the assembly district (AD) and election district (ED) for each ballot. New York City votes are reported in geographically bounded AD/EDs in a manner reminiscent of precinct-level tallies in other states, and in this report we refer to a specific AD/ED combination as a “precinct.” To keep voter choices private, a de-identifying step is taken in which CVRs were associated with numerical identifiers rather than voter names.

Before 2021, if sufficiently few voters cast ballots in a specific precinct, those records were bundled with another precinct. In 2021, this step was not taken. This raises voter privacy concerns, because voter rolls contain precinct information. Therefore election totals with one vote can be matched with voter rolls, raising the possibility of identifying the person associated with an individual vote.

As a byproduct of our in-depth analysis of the mayoral primary election, we confirmed this possibility. **We were able to individually identify 378 voters and their choices in the 2021 New York City mayoral and all lower primary elections.** While this is only a small fraction of all votes cast, it raises a major privacy concern and reveals a conflict between the NYC BOE’s current implementation of reporting requirements and the New York Voter Bill of Rights secret-ballot provision.

## Data and Results

Our analysis used only publicly available data. First, we used the CVR data from the June 2021 primary elections. CVR data must be stored on a single-vote basis in order to do the necessary tabulations for a ranked-choice election. We searched for precincts with only one vote, and found 800 cases.

Next, we turned to the New York State voter file from August 2021, which we obtained by a New York Freedom of Information Law (FOIL) request. This file includes turnout information from past elections, including June 2021. Using the statistical programming language<sup>1</sup> R, we searched the voter file for precincts reporting exactly one voter participating. Taking the intersection with cast vote records, 378 precincts reported both a single voter and a single CVR, 8 for the Democratic primary and 370 for the Republican primary.

In each of these 378 cases, the match therefore contained the voter's name, address, precinct information, and how the single vote was cast. Other information was available in the voter file such as previous voting history, party registration information, and registration status. One match is shown in **Figure 1**, corresponding to the precinct that includes Gracie Mansion, which is home to one registered voter, the current Mayor's son, Dante de Blasio.<sup>2</sup>

**Figure 1: An example of unique cast-vote and voter-file records from the same electoral district (precinct).** Here, Assembly District 76, Election District 89 contains multiple cast-vote records, making re-identification impossible. Election District 90 contains exactly one cast-vote record and one voter listed.

Precinct:					Voter List Data:				
AD: 76 ED: 089					AD: 76 ED: 089				
Cast Vote Rec	vote_1	vote_2	vote_3	vote_4	vote_5	Last Name	First Name	Middle Name	Address
64107	Wiley	Garcia	Yang	undervote	undervote	g			334 EAST 78 STREET NEW YORK, New York 11225
64109	Garcia	Wiley	Yang	undervote	undervote	g			345 EAST 77 STREET NEW YORK, New York 11201
64111	Garcia	undervote	undervote	undervote	undervote	r			341 EAST 77 STREET Brooklyn, New York 11212
64113	Garcia	Adams	Yang	undervote	undervote	...			
64116	Stringer	Garcia	undervote	undervote	undervote				
64117	Garcia	Stringer	Wiley	Donovan	Yang				
64119	Garcia	Yang	Wright Jr.	Donovan	Wiley				
64121	Wiley	Garcia	Adams	Stringer	Wright Jr.				
AD: 76 ED: 090					AD: 76 ED: 090				
1005912	Wiley	Adams	Garcia	McGuire	Donovan	de blasio	DANTE	Q	181 East End Avenue New York, New York 11209
AD: 76 ED: 091					AD: 76 ED: 091				

**Table 1** shows the number of precincts with small numbers of votes. Of the 422 single-vote precincts where an exact match was not found, the voter file either showed no voters (69 precincts) or more than one voter (353 precincts). These discrepancies may have arisen from incomplete reporting of voter turnout from New York City to the state, changes in the voter file between June and August, or discrepancies in the cast-vote records such as precinct misassignment or ballot disqualification. More matches should be possible by incorporating voter files from other dates.

<sup>1</sup> The same comparison could also be done in a spreadsheet program.

<sup>2</sup> The voting patterns of Dante de Blasio have been the subject of reporting since he first registered. Because Gracie Mansion is the only location with an exclusive AD/ED and that specific AD/ED is reported separately, deducing Dante de Blasio's voting record has been possible in past elections.

These discrepancies also raise the possibility that some apparent re-identifications may be a coincidence of multiple discrepancies.

**Table 1: Number of election districts (precincts) with small numbers of votes.**

<b>Number of Votes Cast in Precinct</b>	<b>Republican primary</b>	<b>Democratic primary</b>
1	776	24
1, with voter-file match	370	8
2-9	2348	268
10+	1628	5316

**A simple solution: aggregation of the smallest precincts**

A release of re-identifiable voting data violates New York state law §7–202(j), which stipulates that votes be retained “in a manner intended and designed to protect the privacy of the voter.” On September 13, we brought this to the attention of the New York City BOE through a telephone conversation with a bipartisan delegation of the BOE and their general counsel. They were aware of the re-identification risk and noted that we were the first research team to find it. In their interpretation, they believe that legal reporting requirements to make data available in a particular format does not allow them latitude to group single-voter precincts, as was previously done. They thought that changing this practice would require a change to the New York City Charter.

The privacy loophole we found is easily fixed. Our analysis relied on the certainty of single-voter precincts. If the data reporting requirements were changed to require records from single-vote precincts to be grouped with neighboring precincts, it would no longer be possible to re-identify voters. This operational fix would take a few minutes of staff time per election, and would bring BOE procedures into line with the New York State Voter Bill of Rights.

This solution also maintains the openness of records. Both data sources we used, cast-vote records and the New York state voter file, are integral to evaluating the fairness and administration of elections. Cast-vote records allow citizens to monitor the fair tabulation of elections, and the voter file is necessary to ensure that districts are drawn to ensure fair and equitable representation. Conversely, restricting the voter file would make this information effectively available only to political parties and campaigns, super political action committees (super PACs), and other interest groups.

## **Conclusion**

It is our hope that by raising this issue we may better protect the privacy of voters in New York City in a way that preserves important access to election data. These two interests—voter privacy and data transparency—do not have to come into conflict. By allowing re-aggregation of small precincts, neither principle would have to be sacrificed.

## **About the authors**

Jesse T. Clark ([jc0639@princeton.edu](mailto:jc0639@princeton.edu)) is a Postdoctoral Researcher at the Electoral Innovation Lab, Princeton University. Lindsey Cormack ([lcormack@stevens.edu](mailto:lcormack@stevens.edu)) is an Assistant Professor of Political Science at The Stevens Institute of Technology. Sam Wang ([sswang@princeton.edu](mailto:sswang@princeton.edu)) is Professor of Neuroscience at Princeton University and is Director of the Electoral Innovation Lab.