

Problems with Alexander Holtzman's 2012 Stanford honors senior thesis,  
"The Unanticipated Inequalities of Electoral Reform:  
Racial and Ethnic Disparities in Voting Behavior  
under Oakland's Ranked Choice Voting Program"

[Note: The following concerns have been raised by David Cary. They are not the only concerns, but they include some of the more serious issues that render most of Holtzman's results not credible.]

Holtzman starts with a general hypothesis that an alleged greater complexity associated with voting in a ranked choice voting (RCV) contest causes a decline in voter participation, both in terms of voter turnout and in the effectiveness of voting, and that the impact is greater for traditional racial and ethnic minorities. He contrasts this hypothesis with a claim often made by RCV proponents that RCV increases turnout.

In testing his hypothesis, Holtzman's quantitative analysis has three parts, the first of which purports to be a within-city turnout comparison for Oakland California between its 2006 mayoral contest without RCV and its 2010 mayoral contest with RCV. The 2006 contest was conducted under a two-round runoff system that used a June primary consolidated with the state-wide gubernatorial primary and a contingent November runoff consolidated with the state-wide gubernatorial general election. In 2006 the mayoral contest was won in June and there was no November runoff.

However it appears that Holtzman did not use any turnout data from the June 2006 primary for the within-city comparison. That contest had a reported turnout of 46.0%. The comparable mayoral turnout in 2010 (contest specific, ballots cast / registered voters) is 60.09% (122,268 / 203,469), an increase of 14.1 percentage points (30.6% relative increase). Instead Holtzman appears to use election-wide turnout data from November 2006 and compares that to election-wide turnout data from November 2010. Both of those election turnouts match the numbers he reports: 60.22% for 2006, 61.23% for 2010.

As a result of using data from the wrong election, Holtzman underestimates the increase in turnout by more than an order of magnitude. Using such grossly erroneous data ruins much of his core analysis and results, rendering his conclusions not credible. This is a principle reason why I suggested that he withdraw the paper.

The comparison he appears to actually make would be a good way to test whether gubernatorial turnout increased in Oakland concurrent with the switch to RCV for local contests and how precinct-level gubernatorial turnout might have responded similarly or differently for precincts with different racial/ethnic group mixes. Of course such a comparison is at least twice removed from the hypothesis he started with and from the claims that RCV proponents have primarily made about turnout.

First, the principle claim made by RCV proponents is that compared to a two-round runoff system RCV will improve turnout in the contests where it is used, not necessarily in other contests in the same or different elections. Second, the mechanism that proponents claim will increase turnout is not that RCV will necessarily bring more voters to the polls for a given election, but that a lower turnout election can be avoided, the entire contest can be decided by voters in a single election that is scheduled when the most voters typically vote. The claimed turnout improvement is contest centric, not election centric. As applied to Oakland, this meant that RCV allowed the mayoral contest to avoid the June election, which over the years had consistently lower turnout, and to be fully decided in the consistently higher November election. Holtzman mentions this phenomenon elsewhere in the paper (p. 7) – that a contest decided in two elections will typically be partially decided by one election with lower turnout. However he describes it in terms that are specific to his home town, San Francisco which previously used a two-round runoff system with a November / December schedule and for most contests the December runoff election was the low-turnout election that RCV avoided.

While Holtzman mentions the phenomenon of roll off, in his quantitative analysis he largely ignores any effects related to elections with multiple contests.

Correctly using the June 2006 data instead of November 2010 data would change the base line turnout percentages. One could also reasonably expect that it might change the racial/ethnic mix of the base line turnout, which additionally undermines the credibility of Holtzman's results.

The comparison that Holtzman claims to be making would show a different graph of 2006 data in Figure 3-1 (p. 40), perhaps with the data in a different shape or orientation, but with the data generally shifted downward by about 0.13. Likewise the comparison he claims to be making would show a different Figure 3-2 (p 41), perhaps with the data in a different shape or orientation, but with the data generally shifted upward by about 0.13, much of it "off the chart".

But the problems with this comparison are more extensive. Even given the data that Holtzman did analyze, he appears to misapply and misinterpret the statistical analysis of that data. The best fit line in Figure 3-2 is everywhere positive, meaning that at every "Percent Minority Precinct Population", from 0% to 100%, his quantitative model predicts an increase in turnout. Yet he claims that "while overall turnout increased during the implementation of RCV, minority turnout decreased" (p. 41).

He also appears to apply statistical tests as if the precinct data came from an independent random sampling of precinct data from a larger "true" population. But this does not appear to be the case. The indications are that he is analyzing all or nearly all of the precincts. The justification for his statistical tests is then questionable and his statistical analysis portrays a greater statistical certainty and predictive ability for the overall effects that he generally ascribes to RCV than the comparison of just two elections warrants.

Furthermore, it appears that he tests various coefficients to see if their difference from zero is statistically significant. Yet in cases where statistical significance is indicated, he often seems to misinterpret those results to mean that there was an actual decline in turnout, as opposed to less of an increase.

The possibility that precincts comprised mostly of one racial/ethnic group might have less of an increase in turnout than precincts comprised mostly of another racial/ethnic group is an appropriate possibility to check, but that should be distinguished, in testing and interpretation, from the possibility that the turnout increase was proportionally less or that the turnout actually declined.

One of the things Holtzman does not disclose for this comparison is how much of the variation in turnout change is actually associated with the predictor variables. He highlights that information in other statistical analyses. His selective disclosure, the magnitude of the coefficients, and Figure 3-2 suggests that the predictor variables only explain a small part of that variation. Elsewhere he mentions other factors that could play a role, but he mostly ignores them when developing his quantitative model and interpreting the results.

The second part of the analysis, the difference-within-difference model involving Long Beach, apparently builds on the same wrong Oakland data that was used in the first comparison. It also apparently incorporates gubernatorial election turnout data for Long Beach. No municipal elections for office are involved in the Long Beach data. That may avoid some confounding factors but fails to neutralize others and may introduce more issues than it resolves, something that Holtzman fails to account for. In any case, because of which Oakland data was used, like the first comparison, this comparison might be relevant for a different question, but the conclusions reached are similarly untrustworthy.

The third part of the analysis turns to questions of ballot usage and exhausted ballots, looking only at the Oakland 2010 mayoral data. Holtzman's quantitative analysis and conclusions largely ignore that there are legitimate reasons why a voter who selected plurality front-runner Don Perata as a first choice would feel less likely to vote for a second or third candidate and that this does not necessarily indicate a disenfranchisement of the voter, a failure of RCV, or confusion on the part of the voter.

Much of Holtzman's analysis in the third section of comparisons relies on the faulty premise that under rational choice theory "Having overcome the costs of getting to the polling place, the only rational reason to support only one candidate would be if the voter were truly indifferent among all the other candidate choices." (p. 22) This premise might have some validity when constrained by suitable assumptions, for example if the voter's decision horizon is limited to just a single contest or election. But both individual voters and large groups of coordinated voters are well known for adopting a longer term strategy of withholding support and votes from their most preferred viable candidate or even their most preferred candidate. Such a strategy is entirely rational and is most commonly used, often with great expertise and deliberation, in non-RCV

contests. Such a strategy is also incorporated in heuristics for voting decisions that can be superficially mistaken for being less sophisticated or rational. The rather simplistic model of rational behavior that Holtzman naively uses undermines the validity of the conclusions he reaches.

Just as questions of turnout deserved a comparison to the alternative of a two-round runoff system, so do questions of exhausted votes. The lack of any such comparison greatly limits the possible scope of valid conclusions about RCV and exhausted votes. RCV often performs much better than traditional alternatives when evaluated on the basis of exhausted votes. Because of the flawed conceptual framework of this analysis, the analysis of ballot usage and exhausted votes and the related conclusions are not credible, either.