IL-9 Democratic Primary Simulation Model

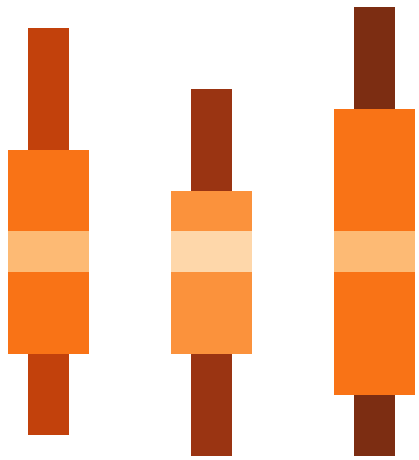
Methodology & Technical Documentation

February 9, 2026 | 100,000 Simulations | 436 Precincts

1. Overview

This model uses a Monte Carlo simulation approach to explore the range of plausible outcomes in the Illinois 9th Congressional District Democratic primary. We run 100,000 simulated elections across all 436 precincts in the district, drawing from probability distributions for turnout, candidate support, and geographic variation. The goal here is not to predict the winner with certainty — that would require better polling data than anyone has for a House primary — but to map out the landscape of what could happen and how likely different scenarios are. Think of it as stress-testing the race rather than calling it.

Each simulation draws a random turnout level, applies geographic and demographic shocks, and computes vote shares for all seven candidates in every precinct. By aggregating across 100,000 runs, we can estimate win probabilities, identify competitive precincts, and understand the conditions under which different candidates are most likely to succeed.



2. Data Sources

Precinct Demographics The model operates at the precinct level across 436 precincts spanning Cook County, the City of Chicago, Lake County, and McHenry County. For each precinct, we compiled 79 demographic and political variables from the American Community Survey (ACS) and 2024 general election results. These include age distributions, racial composition, educational attainment, income levels, housing type, and past partisan voting patterns. The 2024 presidential results at the precinct level give us a recent read on the political baseline.

Polling Data
We incorporated data from six polls of the race. However, only one of these — the Data for Progress (DFP) survey — is truly independent. The rest are campaign internals, which we discount significantly in the model. Campaign polls tend to oversample favorable demographics and frame questions in ways that flatter the sponsoring candidate. A notable example: Fine's internal poll showed her at 21%, while independent polling consistently placed her in the 8-10% range. We weight the DFP poll much more heavily and use the internals mainly as directional signals rather than point estimates.

DFP Crosstabs
The DFP survey (N=569) provided crosstabs breaking down vote choice by age group, ideological self-identification, gender, and education level. These crosstabs are the backbone of our candidate affinity model — they tell us, for example, how much more likely a voter under 40 is to support Abughazaleh versus Biss, or how ideology maps onto candidate preference. We use these relationships to translate precinct demographics into baseline vote shares.

3. Model Architecture

Turnout Model

District-wide turnout is drawn from a normal distribution with mean 150,000 and standard deviation 22,000, truncated at 80,000 and 240,000 to avoid implausible extremes. This range reflects the genuine uncertainty in primary turnout — House primaries can swing wildly depending on whether there's a competitive presidential primary on the same ballot, weather, and general enthusiasm. Precinct-level turnout is allocated using primary-to-general ratios derived from 2022 Cook County Assessor data. For Lake and McHenry counties, where we have less historical primary data, we use OLS regression on available demographic predictors to estimate primary participation rates.

Candidate Priors

Each candidate's baseline vote share in a precinct is built from a 50/50 blend of age affinity and ideology affinity scores, derived from the DFP crosstabs. For example, a precinct with lots of young progressive voters will start with a higher prior for Abughazaleh, while a precinct skewing older and more moderate will lean toward Biss or Fine. These raw affinity scores are then rescaled so that the district-wide average matches the polling targets. Undecided voters are allocated conservatively — we don't assume they'll break proportionally for frontrunners. Instead, we distribute them with a slight lean toward less-known candidates on the theory that undecideds in a primary are often looking for alternatives.

Candidate-Specific Adjustments

On top of the demographic baseline, we apply several candidate-specific adjustments that reflect real-world factors the demographics alone don't capture:

- **Biss (Evanston boost)**: Biss gets a 1.4x multiplier in Evanston precincts, reflecting his home-turf advantage as the former state senator from the area. Additionally, the Schakowsky endorsement adds +2.5 percentage points, weighted by the share of older voters in each precinct (since Schakowsky's endorsement resonates more with longtime district residents).
- **Abughazaleh (Chicago youth boost)**: Kat receives an upward adjustment in Chicago precincts proportional to the share of 20-39 year-old residents. This captures her strength among younger, more progressive urban voters who are less likely to show up in traditional polling.
- **Andrew (North Shore boost)**: Andrew gets a 2.2x multiplier in New Trier and Niles township precincts, reflecting his deep local ties and name recognition in the North Shore communities. This is why his average vote share is a respectable 8.2% despite near-zero win probability — he runs up big numbers in a handful of precincts but has minimal support elsewhere.
- **Evanston undecided redirect**: In Evanston precincts, a portion of the undecided vote is redirected away from Biss. The logic is that voters in his home base who are still undecided probably aren't natural Biss supporters — if they were, they'd already be with him.
- **Fine (Glenview boost)**: Andrew gets a 1.4x multiplier in Glenview township precincts, reflecting his deep local ties and name recognition in the communities that she bith represnts and lives.

5. Simulation Engine

Each of the 100,000 simulated elections proceeds through several stochastic layers that inject realistic uncertainty into the results:

Geographic Shocks

We apply region-specific random shocks to capture the reality that turnout and enthusiasm vary by geography in ways that polls don't fully measure. Chicago precincts get a shock drawn from N(0, 0.12), suburban Cook precincts from N(0, 0.08), and Lake/McHenry precincts from N(0, 0.10). These shocks are multiplicative — a positive shock in Chicago means all Chicago precincts see slightly higher turnout and potentially different candidate dynamics in that particular simulation.

Turnout Composition Effects

The model captures a key dynamic in primaries: high-turnout elections tend to favor establishment candidates, while low-turnout elections favor progressives and insurgents. When the turnout draw is above the mean, Biss and Fine get a small boost; when it's below the mean, Abughazaleh and other progressive candidates benefit. This isn't arbitrary — it reflects the well-documented pattern that occasional voters who only show up in high-turnout primaries tend to default to better-known candidates.

Candidate Lane Correlations

Candidates don't exist in isolation — they compete within ideological lanes. We model two correlated shock terms: a progressive lane shock (standard deviation = 4) that simultaneously affects Abughazaleh, Amiwala, Simmons, and Huynh, and an establishment lane shock (standard deviation = 4) that affects Biss and Fine together. When something boosts one progressive candidate, it tends to boost all of them (and vice versa). This captures dynamics like a progressive endorsement that energizes the whole lane, or a news cycle that benefits the establishment wing broadly.

Polling Error and Noise

We add a systematic polling error term (standard deviation = 5 percentage points) that shifts all candidates uniformly in a given simulation. This accounts for the possibility that the polls are systematically off in one direction — maybe Biss is actually 5 points weaker than the polls suggest, or maybe he's 5 points stronger. On top of that, each precinct gets its own idiosyncratic noise term to reflect the fact that individual precincts can behave differently from what their demographics would predict.

6. Limitations

We want to be upfront about what this model can't do. There are several meaningful limitations to keep in mind:

- **Limited independent polling**: With only one truly independent poll (DFP, N=569), we're building on a thin empirical foundation. The margin of error on a 569-person survey is already around ±4%, and crosstab subgroups are even noisier. If that one poll is off, our baseline priors are off too.
- **No campaign finance data**: We don't incorporate fundraising totals, spending patterns, or advertising buys. In a primary where name recognition is everything, a candidate who can afford heavy mail and digital advertising might outperform their polling baseline. This is a real gap.
- **Endorsement effects are estimated**: The Schakowsky endorsement boost for Biss is our best guess based on analogous races, but we don't have experimental evidence for how much a retiring incumbent's endorsement moves votes in this specific district. The true effect could be larger or smaller.
- **Demographic proxies are imperfect**: We're using Census demographics as proxies for primary voter behavior, but primary electorates are not Census populations. Primary voters skew older, more educated, and more politically engaged than the general population. Our precinct-level demographic data describes residents, not likely primary voters.
- **No late-breaking dynamics**: The model is a snapshot. It can't account for an October surprise, a viral debate moment, a major endorsement we don't know about yet, or a scandal. The closer we get to election day, the more real-world events will matter relative to any model.

7. Updates & Contact

This model is updated periodically as new data becomes available — additional polls, updated demographic estimates, or refined assumptions. The current version reflects data as of February 9, 2026. Previous versions may differ in methodology or results as the model evolves.

For questions, feedback, or data requests, reach out on Twitter/X at @bayes_pr or visit il9.org for the latest maps, data, and analysis.