

## Inferences from the DNC Provisional Ballot Voter Survey

Walter R. Mebane, Jr.

April 27, 2005

The survey conducted in Cuyahoga County, Ohio,<sup>1</sup> shows that the single most important cause of voters casting a provisional ballot in the county in the November 2004 election was residential mobility. About 60 percent of the provisional ballots were cast by those who either were voting in Ohio for the first time or who had previously voted in Ohio but had since moved. Among those who had previously voted in Ohio and not moved since doing so, voters younger than 55 years of age were much more likely to cast a provisional ballot than older voters were. Among those who had previously voted in Ohio but since moved, African American voters were more likely than white voters were to cast a provisional ballot

Before considering the detailed results from the survey, note that the matched, case control design of the Cuyahoga survey means that when the appropriate sampling weights are used, the survey exactly estimates the overall frequency of casting a provisional ballot. From administrative records we know that the proportion of provisional ballots among all ballots cast is 0.03518, a value the survey estimate is constrained to reproduce exactly.<sup>2</sup> Estimates for the proportion of provisional ballots cast by groups of voters in Cuyahoga County will not be exact but will be subject to sampling error. A technical appendix to this memo describes the methods used to compute estimates from the survey data.

Despite the enforced accuracy of the overall proportion, the sample was implemented using a zipcode-level matching design that may introduce bias in the estimates of other quantities when making inferences about all voters in Cuyahoga County. A nonprovisional voter has a positive probability of being included in the second sample only if the voter lives in the same zipcode as a provisional voter. If there are zipcodes in which there are nonprovisional voters but no provisionals, then the nonprovisional voters in those zipcodes have zero probability of being included in the nonprovisional survey.<sup>3</sup> In this case estimates using the nonprovisional survey data are biased. Given the way the estimators for the survey proportions are derived, it may be reasonable to say the survey estimates are biased only if there are zipcodes in Cuyahoga County where it is *impossible* that any provisional ballots were cast.<sup>4</sup> Because the personal experiences and administrative problems that cause a provisional ballot may affect almost any voter (e.g., moving residences, misdirected absentee ballots, record keeping errors), it may be reasonable to rule out this source of bias. The estimates discussed in this report ignore this potential bias.

Tables 1, 2 and 3 show how often Cuyahoga voters cast provisional ballots given various personal attributes and election-day experiences they had. The first column of each table lists

---

<sup>1</sup>DNC Ohio Election Investigation Project Provisional Ballot Survey, Phase 1 and Phase 2, conducted by The Feldman Group and brilliant corners Research & Strategies. Phase 1 interviews were conducted March 2–3, 2005. Phase 2 interviews were conducted March 22–24, 2005.

<sup>2</sup>The administrative estimate is  $0.03518049 = 24463 / (687260 + 8097)$ , where 24,463 is the “Total” from the file `rejected & ok provisionals 04.pdf`, 687,260 is “BALLOTS CAST TOTAL” from the file `110204_GE_Canvass.txt`, (RUN DATE:02/28/05 11:45 AM), and 8097 is the total number of all rejected provisional ballots from file `rejected & ok provisionals 04.pdf`.

<sup>3</sup>The sample of provisional voters includes voters from every residential five-digit zipcode area in Cuyahoga County. Nonprovisional voters were matched to provisional voters in more local domains, however, all the way down to some matches done in zip-9 areas. See the technical appendix for more details.

<sup>4</sup>I have in mind a superpopulation justification for the sample estimators.

Table 1: Provisional Ballots and Attributes Related to Residential Mobility

Attributes	Percent Provisional	Confidence (+ or -)	Percentage of Voters
Ever Voted in Ohio? Yes	3.1	0.2	89.6
Ever Voted in Ohio? No	9.4	7.1	7.3
Moved Since Last Voted in Ohio? Yes	11.4	4.4	15.0 <sup>a</sup>
Moved Since Last Voted in Ohio? No	1.7	0.2	84.2 <sup>a</sup>
Voted in Past at Polling Place	1.7	0.2	77.7
No Vote in Past at Polling Place	11.2	3.9	19.4
At Current Address: Five Years or Less	9.1	2.7	28.2
At Current Address: More than Five Years	1.3	0.2	67.3
Own Home	2.7	0.4	66.2
Rent Home	5.7	1.5	26.8
Registered to Vote in 2004	26.5	12.2	3.0
Already Registered	2.5	0.2	96.1

Notes: <sup>a</sup> Percentage of the 89.6% of voters who previously voted in Ohio.

various attributes that describe different voters. The second column shows the percentage of those voters estimated to have cast a provisional ballot, and the third column shows the range for a 95% confidence interval around the percentage estimate: we can be 95% confident that the true percentage falls within an interval defined as the estimated percentage plus or minus the reported confidence range. The final column shows the percentage of voters in Cuyahoga County who have the indicated attribute.

Why did voters cast provisional ballots? The single most important cause seems to be residential mobility. The first entry in Table 1 shows that of those who said they had previously voted in Ohio, 3.1 percent cast a provisional ballot, while 9.4 percent of those who said they had never voted in Ohio cast a provisional ballot. Because the number of people who said they had never voted in Ohio is small—they are only 7.3 percent of voters—the uncertainty in the estimated percentage of them who cast a provisional ballot is large. The 95% confidence interval ranges from 2.1 percent to 16.7 percent. The estimated percentage casting a provisional ballot among those who had previously voted in Ohio has very little uncertainty, however, so it is much more likely than not that the percentage is substantially smaller in this group.

The message about mobility is further conveyed by the estimated percentage casting a provisional ballot among those who said they had previously voted in Ohio but had moved since the last time they voted. Fifteen percent of those who said they had previously voted in Ohio said they had since moved. Among those movers, 11.4 percent cast a provisional ballot, compared to just 1.7 percent of those who said they had not moved since the last time they voted. The estimated percentage casting a provisional ballot among the movers is again somewhat uncertain. The 95% confidence interval ranges from 7.0 percent to 15.8 percent. But the percentage is clearly larger than the percentage among those who did not move.

If we focus on the point estimates for the percentage casting a provisional ballot among those

who said they had not previously voted in Ohio and those who said they had moved since they last voted in Ohio, it appears that more than 60 percent of the provisional ballots can be accounted for. If 9.4 percent of the first-time voters cast a provisional ballot and 11.4 of those who had moved cast a provisional ballot, that would imply that at least 2.2 percent of voters overall cast a provisional ballot.<sup>5</sup> Recall that among all voters in Cuyahoga County, 3.5 percent cast a provisional ballot. The provisionals associated with the foregoing measures of residential mobility would therefore account for 63 percent of all provisional ballots in the county.

Another indicator supporting the importance of residential mobility in explaining why voters cast provisional ballots is the percentage casting a provisional ballot among those who said they had not previously voted at the polling place where they voted in November 2004. Of the 19.4 percent of voters who said they had not previously voted at the November 2004 polling place, 11.2 percent are estimated to have cast a provisional ballot. The 95% confidence interval for that estimate ranges from 7.3 percent to 15.1 percent. Only 1.7 percent of those who said they had previously voted at the November 2004 polling place cast a provisional ballot. Subtracting the 7.3 percent of voters who said they had not previously voted in Ohio from the 19.4 percent who said they had not previously voted in the November polling place leaves 12.1 percent of voters having previously voted in Ohio but not in the November polling place. That number is close to the percentage of voters who said they had previously voted in Ohio but had since moved, which is 13.4 percent. The excess in the latter number may be accounted for by people who moved but remained within the same polling place boundaries.

Yet more indications of the importance of residential mobility come from other questions that relate to the permanence of each voter's residence. An estimated 9.1 percent of those who said they have lived at their current address less than five years cast a provisional ballot, but only 1.3 percent of those who have lived at their current residence more than five years cast a provisional ballot. Of those who said they rent their home, 5.7 percent are estimated to have cast a provisional ballot, but only 2.7 percent of those who own their home cast a provisional ballot.

Probably not all the voters who said they had never voted in Ohio were new Ohio residents. It is difficult to separate those who moved from those who were already resident in Ohio but were newly mobilized to vote in the 2004 election. Of those who said they were not registered to vote in their county before the presidential election and registered in order to vote in it, an estimated 26.5 percent cast a provisional ballot. Because only three percent of voters said they had newly registered in that way, the 95% confidence interval for that estimate is large, ranging from 14.3 percent to 38.7 percent. But even the lower bound of the interval is vastly higher than the estimated 2.5 percent (plus or minus 0.2 percent) of those who said they were already registered who cast a provisional ballot. Those who newly registered to vote in November 2004 are an undetermined mix of movers and newly mobilized voters.

Table 2 shows that the actions people took when voting were somewhat related to their chances of casting a provisional ballot, but the experience they had when trying to vote was strongly related to those chances. Of those who had requested an absentee ballot, 2.4 percent (plus or minus 1.0 percent) cast a provisional ballot, a number significantly lower than the percentage (3.4 percent, plus or minus 0.3 percent) casting a provisional ballot among those who did not request an absentee ballot. Neither voters' presidential candidate choices nor their reported partisan tendencies were associated with casting a provisional ballot. Among Kerry

---

<sup>5</sup>The calculation is  $.094(7.3) + .114(.15)(89.6) = 2.21836$ .

Table 2: Provisional Ballots and Election Experiences

Attributes	Percent Provisional	Confidence (+ or -)	Percentage of Voters
Requested Absentee Ballot	2.4	1.0	17.2
No Absentee Ballot	3.7	0.3	82.4
Voted for Kerry	4.2	0.6	51.4
Voted for Bush	4.1	1.4	23.3
Party Identification: Democrat	3.7	0.6	53.3
Party Identification: Republican	4.9	1.9	16.8
Party Identification: Other	2.5	1.0	16.6
Identification Request: Had ID	7.9	1.6	31.5
Identification Request: Not Asked	1.4	0.3	59.1
Registration Questioned	19.2	4.6	9.8
Registration Not Questioned	1.9	0.2	80.1

voters, 4.2 percent cast a provisional ballot, and among Bush voters 4.1 percent cast a provisional ballot. There is no statistically significant difference between the percentage of Democratic party identifiers who cast a provisional ballot and the percentage of Republican party identifiers who did so.

Voters who had their right to vote questioned when they appeared at the polls were much more likely to end up casting a provisional ballot. 7.9 percent (plus or minus 1.6 percent) of those who were asked to provide identification at the registration table at their polling place cast a provisional ballot, compared to 1.4 percent (plus or minus 0.3 percent) of those who were not asked for identification. Of those who had their registration questioned at the registration table, 19.2 percent (plus or minus 4.6 percent) cast a provisional ballot, versus 1.9 percent (plus or minus 0.2 percent) of those who did not have their registration questioned. Most likely these percentages do not measure causes for someone to cast a provisional ballot, but instead they represent part of the process of being directed to cast such a ballot.

Table 3 shows that some but not all of the personal attributes of voters that are often found to be related to political activity are related to the frequency of casting provisional ballots. Substantial differences in the frequency of casting a provisional ballot occur across age groups. Older voters are less likely to cast a provisional ballot. Eleven percent (plus or minus 5.1 percent) of voters 34 years of age or younger cast a provisional ballot, while only 5.3 percent (plus or minus 1.6 percent) of voters aged 35–55 years did so. Only 1.6 percent (plus or minus 0.4 percent) of voters aged 55 years or older cast a provisional ballot.

On the whole, voters who identify themselves as African American may be more likely to cast a provisional ballot than are voters who identify themselves as white, but the difference between the two groups is not statistically significant. 3.5 percent (plus or minus 0.5 percent) of white voters cast a provisional ballot, while 4.1 percent (plus or minus 1.0 percent) of African American voters did so. The uncertainty in the estimated percentage for African Americans is such that the 95% confidence interval for that estimate includes most of the 95% confidence interval of the estimated percentage for white voters, but the upper bound of the 95% confidence interval for

Table 3: Provisional Ballots and Personal Attributes

Attributes	Percent Provisional	Confidence (+ or -)	Percentage of Voters
Age: 34 Years or Younger	11.0	5.1	8.7
Age: 35-55 Years	5.3	1.6	28.3
Age: 55 Years or Older	1.6	0.4	52.1
White	3.5	0.5	54.8
African American	4.1	1.0	29.7
Male	3.2	0.7	48.7
Female	3.8	0.5	51.3
Voter is Union Member	3.0	1.4	10.1
Union Member in Household	3.8	2.7	5.9
No Union Member in Household	3.7	0.4	76.9
Education: High School or Less	3.4	0.7	43.5
Education: Some College	3.9	1.2	23.2
Education: College Graduate	4.4	1.5	18.6
Education: Post-graduate	2.9	1.8	9.3
Marital Status: Married	3.8	0.8	38.2
Marital Status: Not Married	3.6	0.5	55.5
Income: \$20K or Less	4.4	1.3	17.8
Income: \$20-40K	3.8	1.2	22.3
Income: \$40-60K	4.9	2.8	10.2
Income: \$60K or More	5.1	1.8	16.0
Income: Not Ascertained	1.7	0.5	33.7

white voters is less than the point estimate for African American voters. The estimated difference between the percentages (0.5) is slightly smaller than the estimated standard error of the difference (0.6).

Likewise, female voters may be more likely to cast a provisional ballot than male voters are, but the statistical significance of the difference is questionable. 3.8 percent (plus or minus 0.5 percent) of female voters cast a provisional ballot, while only 3.2 percent (plus or minus 0.7 percent) of male voters did so. The estimated difference between the percentages (0.5) is slightly larger than the estimated standard error of the difference (0.4).

Only one other statistically significant difference in the percentage casting provisional ballots occurs across the other groups shown in Table 3. There are no statistically significant differences associated with union membership, education, marital status or measured income.<sup>6</sup> Those for whom income was not ascertained were much less likely to cast a provisional ballot. Among those voters, 1.7 percent (plus or minus 0.5 percent) cast a provisional ballot, while among voters who responded to the income survey item the estimates range from 3.8 percent to 5.1 percent

<sup>6</sup>Within the group of unmarried voters, those who said they were widowed had a substantially lower frequency (1.4 percent, plus or minus 0.5 percent) of casting a provisional ballot. Presumably this difference is a reflection of older voters being less likely to cast a provisional ballot.

Table 4: Provisional Ballots, Personal Attributes and Residential Mobility

Attributes	Percent Provisional	Confidence (+ or -)
Ever Voted in Ohio? Yes: White	3.0	0.4
Ever Voted in Ohio? Yes: African American	3.9	1.0
Moved? Yes: White	8.8	4.4
Moved? Yes: African American	16.2	9.1
Moved? No: White	1.6	0.4
Moved? No: African American	2.1	0.7
Ever Voted in Ohio? Yes: Age 34 or Under	10.3	4.6
Ever Voted in Ohio? Yes: Age 35–54	5.1	1.6
Ever Voted in Ohio? Yes: Age 55 or Over	1.5	0.4
Moved? Yes: Age 34 or Under	19.6	13.1
Moved? Yes: Age 35–54	9.4	6.3
Moved? Yes: Age 55 or Over	9.8	5.1
Moved? No: Age 34 or Under	5.2	3.5
Moved? No: Age 35–54	3.4	1.3
Moved? No: Age 55 or Over	0.8	0.3

casting a provisional ballot.

Table 4 shows that the difference between African American voters and white voters in the probability of casting a provisional ballot becomes statistically significant when residential mobility is taken into account. When only voters who said they had previously voted in Ohio are considered, 3.9 percent (plus or minus 1.0 percent) of African American voters cast a provisional ballot while 3.0 percent (plus or minus 0.4) percent of white voters did so. When these previous Ohio voters are separated into those who said they had moved since the last time they voted and those who said they had not moved, the percent casting a provisional ballot is 16.2 (plus or minus 9.1 percent) among African American movers, 8.8 (plus or minus 4.4 percent) among white movers, 2.1 (plus or minus 0.7 percent) among African American nonmovers and 1.6 (plus or minus 0.4 percent) among white nonmovers. For all of these percentages the point estimate among African American voters is greater than the corresponding point estimate among white voters, and the estimated differences between the percentages (0.9, 7.4 and 0.5) are larger than their estimated standard errors (0.5, 5.2 and 0.4). But only the first two differences are statistically significant, at a 90 percent level (one-tailed). The sharpest thing to say is that there is a significant difference between African Americans and whites who previously voted but since moved, but not between African Americans and whites who previously voted but did not move. Comparing the point estimates between African Americans who moved and whites who moved shows the extra risk for African Americans of casting a provisional ballot is substantial. Among those who previously voted in Ohio but since moved, African American voters are 1.8 times more likely to cast a provisional ballot than white voters are.

In contrast, Table 4 shows there is no statistically significant difference in the probability of casting a provisional ballot between the oldest voters and middle-aged voters who moved since

the last time they voted in Ohio. Among movers, 9.4 percent (plus or minus 6.3 percent) of voters aged 35–54 cast a provisional ballot, while 9.8 percent (plus or minus 5.1 percent) of voters 55 or older did so. The 95% confidence intervals for these estimates substantially overlap. 19.6 percent (plus or minus 13.1 percent) of voters younger than 35 who moved cast a provisional ballot. The 95% confidence interval for this youngest group includes the point estimates for the older voters, but the upper bound of the 95% confidence intervals for the older voters is less than the point estimate for the youngest voters. The differences between the youngest group of movers and the two older groups are statistically significant at a 90 percent level (one-tailed). The difference between the oldest voters and the others persists in the group of voters who previously voted in Ohio and did not move since then. Among nonmovers, 5.2 percent (plus or minus 3.5 percent) of voters younger than 35 and 3.4 percent (plus or minus 1.3 percent) of voters aged 35–54 cast a provisional ballot, while 0.8 percent (plus or minus 0.3 percent) of voters 55 or older did so.

## Technical Appendix

The survey consists of two samples from Cuyahoga County, one of voters who cast provisional ballots (henceforth “provisionals”) and one of voters who did not. The provisionals are sampled from a list that contains all people who cast a provisional ballot. The voters are sampled from voters in the same local areas as the provisionals included in that sample: one voter is sampled from a local area for every provisional included in the sample who is from that local area. The local areas in practice are zip-code areas. Nonprovisional voter matches were found in 12 zip-9 areas, 135 zip-8 areas, 224 zip-7 areas, eight zip-6 areas, seven zip-5 areas and two zip-4 areas. For brevity I refer to each area simply as a zipcode. In thirteen instances two nonprovisional voters were sampled from the same zipcode. Each sample has size  $n = 400$ .

The goal is to estimate the conditional distribution of casting a provisional ballot with respect to various variables that were measured in the survey instruments used with each sample. This note sketches how to do that, given the matched sample design.

Let  $y$  denote the event “cast a provisional ballot” and let  $\neg y$  denote the event “cast a nonprovisional ballot.” Let  $p_y$  denote the probability of  $y$  and let  $p_{y|x}$  denote the conditional probability of  $y$  given  $x$ . Let  $N$  denote the total number of voters turning out to vote in Cuyahoga County, with  $N_y$  being the number of provisionals and  $N_{\neg y}$  being the number of nonprovisionals,  $N = N_y + N_{\neg y}$ . The sample of provisionals is  $S_y$  and the sample of nonprovisionals is  $S_{\neg y}$ , with respective sample sizes  $n_y$  and  $n_{\neg y}$ . The probability that individuals are included in the nonprovisional sample varies from person to person because for the matched sample the number of voters in each zipcode varies. The inclusion probability for provisional voter  $i$  is denoted  $\pi_i$  and the inclusion probability for voter  $i$  in the matched sample is denoted  $\nu_i$ . For the provisional sample I assume simple random sampling without replacement, ignoring any phone number matching problems and complications related to getting respondents on the phone. Given the sample sizes  $n_y = n_{\neg y} = 400$  and assuming equal probabilities of inclusion for all provisionals, the sampling fraction from the provisional population is  $\pi_i = f_y = n_y/N_y$ .

I make the simplifying assumption that everyone in the first sample cast a provisional ballot while no one in the second sample did. In fact, some respondents in the first sample deny that they cast a provisional ballot. They may be mistaken, or the telephone methodology may have reached the wrong person. The assumption about voters in the second sample seems to be close to correct.

For various attributes  $x$ , the goal is to estimate the proportion of voters who have  $x$  who cast a provisional ballot. Using  $N_{xy}$  to denote the number of voters who have  $x$  who cast a provisional ballot and  $N_{x\neg y}$  to denote the number of voters who have  $x$  who did not cast a provisional ballot and  $N_x = N_{xy} + N_{x\neg y}$  to denote the number of voters with  $x$ , the proportion of interest may be written  $p_{y|x} = N_{xy}/N_x$ . Neither  $N_{xy}$ ,  $N_{x\neg y}$  nor  $N_x$  is known, but if  $x$  is measured in the surveys the totals may be estimated using the survey data. Using  $x_{yi} = 1$  if provisional voter  $i$  has attribute  $x$  and  $x_{yi} = 0$  if not, Horvitz-Thompson estimators (stabilized for  $\tilde{N}_{x\neg y}$ ) for the totals may be written

$$\tilde{N}_{xy} = \sum_{i=1}^{n_y} \frac{x_{yi}}{\pi_i}, \quad \tilde{N}_{x\neg y} = N_{\neg y} \left( \sum_{j=1}^{n_{\neg y}} \frac{1}{\nu_j} \right)^{-1} \sum_{j=1}^{n_{\neg y}} \frac{x_{\neg yj}}{\nu_j}, \quad \tilde{N}_x = \tilde{N}_{xy} + \tilde{N}_{x\neg y}.$$

A natural estimator for the proportion of interest is then

$$\tilde{p}_{y|x} = \tilde{N}_{xy} / (\tilde{N}_{xy} + \tilde{N}_{x\neg y}).$$



Likewise, the estimator for  $p_{-y|x}$  is

$$\tilde{p}_{-y|x} = \tilde{N}_{x-y} / (\tilde{N}_{xy} + \tilde{N}_{x-y}) = 1 - \tilde{p}_{y|x}.$$

A straightforward way to estimate the sampling variance of  $\tilde{p}_{y|x}$  is to use Taylor linearization (Särndal, Swensson and Wretman 1992, 172–181; Wolter 1985, 221–225). First obtain the partial derivatives of  $\tilde{p}_{y|x}$  with respect to  $\tilde{N}_{xy}$  and  $\tilde{N}_{x-y}$ :

$$\frac{\partial \tilde{p}_{y|x}}{\partial \tilde{N}_{xy}} = \frac{\tilde{N}_{x-y}}{(\tilde{N}_{xy} + \tilde{N}_{x-y})^2}; \quad \frac{\partial \tilde{p}_{y|x}}{\partial \tilde{N}_{x-y}} = -\frac{\tilde{N}_{xy}}{(\tilde{N}_{xy} + \tilde{N}_{x-y})^2},$$

so that given estimators  $\hat{V}(\tilde{N}_{xy})$  and  $\hat{V}(\tilde{N}_{x-y})$  for the variance of each estimated total and a estimator  $\widehat{cov}(\tilde{N}_{xy}, \tilde{N}_{x-y})$  for their covariance, a variance estimator for  $\tilde{p}_{y|x}$  is

$$\begin{aligned} \hat{V}(\tilde{p}_{y|x}) &= \left( \frac{1 - \tilde{p}_{y|x}}{\tilde{N}_{xy} + \tilde{N}_{x-y}} \right)^2 \hat{V}(\tilde{N}_{xy}) + \left( \frac{\tilde{p}_{y|x}}{\tilde{N}_{xy} + \tilde{N}_{x-y}} \right)^2 \hat{V}(\tilde{N}_{x-y}) \\ &\quad - 2 \frac{\tilde{p}_{y|x}(1 - \tilde{p}_{y|x})}{(\tilde{N}_{xy} + \tilde{N}_{x-y})^2} \widehat{cov}(\tilde{N}_{xy}, \tilde{N}_{x-y}). \end{aligned}$$

Assuming simple random sampling among the provisionals we have

$$\hat{V}(\tilde{N}_{xy}) = N_y^2 \frac{1 - f_y}{n_y} \hat{s}_{yx}^2, \quad \hat{s}_{yx}^2 = \sum_{i=1}^{n_y} \frac{(x_{yi} - \tilde{N}_{xy}/N_y)^2}{n_y - 1}.$$

The fact that the sample of nonprovisionals is matched to the realized sample of provisionals makes derivation of  $\hat{V}(\tilde{N}_{x-y})$  and  $\widehat{cov}(\tilde{N}_{xy}, \tilde{N}_{x-y})$  more complicated.

The probability that  $j \in S_{-y}$ , denoted  $\nu_j$ , depends on  $S_y$ . If  $j$  is any nonprovisional voter, then  $j \notin S_{-y}$  if  $j$  is not in the same zipcode as a provisional voter  $i \in S_y$ . Given  $S_y$ , the probability that  $j \in S_{-y}$  depends on the number of voters in the zipcode of the corresponding provisional voter  $i \in S_y$ . Let  $\nu_{j|S_y}$  denote that conditional probability. Let  $d_j$  denote the zipcode of voter  $j$ , and let  $N_{-y d_j}$  denote the number of voters in that zipcode. If there are  $n_{y d_j}$  provisionals in  $S_y$  from zipcode  $d_j$ , then

$$\nu_{j|S_y} = \frac{n_{y d_j}}{N_{-y d_j}}.$$

Using the fact that each  $j \in S_{-y}$  is matched to one  $i \in S_y$ , we have

$$\nu_j = \frac{f_y}{N_{-y d_j}}.$$

Now using the joint inclusion probabilities<sup>7</sup>

$$\nu_{jk} = \begin{cases} \nu_j \nu_k & \text{if } j \neq k, \\ \nu_j & \text{if } j = k, \end{cases}$$

---

<sup>7</sup>This definition of the joint inclusion probabilities ignores the complications associated with the 13 instances where two nonprovisionals were in the same zipcode, as well as the complications due to matches being done in nested zipcode areas (e.g., one match occurring in zip-8 area 44040960 and another in zip-5 area 44040). Ignoring the nested zipcodes, which occur frequently in the sample, probably means the computed estimate  $\widehat{cov}(\tilde{N}_{xy}, \tilde{N}_{x-y})$  is too small. Because that covariance is usually positive, the reported sampling variances may be slightly too large.

we have the stabilized estimator

$$\hat{V}(\tilde{N}_{x^{-y}}) = N_{^{-y}}^2 \left( \sum_{j=1}^{n_{^{-y}}} \frac{1}{\nu_j} \right)^{-2} \sum_{j=1}^{n_{^{-y}}} (1 - \nu_j) \left( \frac{x_{^{-y}j}}{\nu_j} \right)^2.$$

If  $j$  is matched to  $i$  then the joint inclusion probability for  $i$  and  $j$  is  $\nu_j$ , otherwise it is  $f_y \nu_j$ . Let  $i : j$  indicate the  $i$  that is matched to  $j$ . The stabilized covariance estimator is

$$\widehat{cov}(\tilde{N}_{xy}, \tilde{N}_{x^{-y}}) = N_{^{-y}} \left( \sum_{j=1}^{n_{^{-y}}} \frac{1}{\nu_j} \right)^{-1} \sum_{j=1}^{n_{^{-y}}} (1 - \pi_{i:j}) \frac{x_{yi:j} x_{^{-y}j}}{\pi_{i:j} \nu_j}$$

The variance estimator for  $\tilde{p}_{y|x}$  is

$$\begin{aligned} \hat{V}(\tilde{p}_{y|x}) &= \left( \frac{1 - \tilde{p}_{y|x}}{\tilde{N}_{xy} + \tilde{N}_{x^{-y}}} \right)^2 N_y^2 \frac{1 - f_y}{n_y} \hat{S}_{yx}^2 \\ &+ \left( \frac{\tilde{p}_{y|x}}{\tilde{N}_{xy} + \tilde{N}_{x^{-y}}} \right)^2 N_{^{-y}}^2 \left( \sum_{j=1}^{n_{^{-y}}} \frac{1}{\nu_j} \right)^{-2} \sum_{j=1}^{n_{^{-y}}} (1 - \nu_j) \left( \frac{x_{^{-y}j}}{\nu_j} \right)^2 \\ &- 2 \frac{\tilde{p}_{y|x}(1 - \tilde{p}_{y|x})}{(\tilde{N}_{xy} + \tilde{N}_{x^{-y}})^2} N_{^{-y}} \left( \sum_{j=1}^{n_{^{-y}}} \frac{1}{\nu_j} \right)^{-1} \sum_{j=1}^{n_{^{-y}}} (1 - \pi_{i:j}) \frac{x_{yi:j} x_{^{-y}j}}{\pi_{i:j} \nu_j} \\ &= \frac{1}{(\tilde{N}_{xy} + \tilde{N}_{x^{-y}})^2} \left[ \frac{(1 - \tilde{p}_{y|x})^2 N_y^2 (1 - f_y)}{n_y} \sum_{i=1}^{n_y} \frac{(x_{yi} - \tilde{N}_{xy}/N_y)^2}{n_y - 1} \right. \\ &+ \tilde{p}_{y|x}^2 N_{^{-y}}^2 \left( \sum_{j=1}^{n_{^{-y}}} \frac{1}{\nu_j} \right)^{-2} \sum_{j=1}^{n_{^{-y}}} (1 - \nu_j) \left( \frac{x_{^{-y}j}}{\nu_j} \right)^2 \\ &\left. - \frac{2\tilde{p}_{y|x}(1 - \tilde{p}_{y|x})(1 - f_y)}{f_y} N_{^{-y}} \left( \sum_{j=1}^{n_{^{-y}}} \frac{1}{\nu_j} \right)^{-1} \sum_{j=1}^{n_{^{-y}}} \frac{x_{yi:j} x_{^{-y}j}}{\nu_j} \right]. \end{aligned}$$

Values for the constants used in the estimators are in Table 5.

Table 5: Useful Constants

$N$	$695357 = 687260 + 8097$
$N_y$	24463
$N_{-y}$	$670894 = 687260 - 16366$
$f_y$	$0.016351224 = 400/24463$

Sources:  $N$ , 687260 is “BALLOTS CAST TOTAL” from 110204\_GE\_Canvass.txt, (RUN DATE:02/28/05 11:45 AM), and 8,097 is the total number of all rejected provisional ballots from file rejected & ok provisionals 04.pdf;  $N_y$ , “Total” from file rejected & ok provisionals 04.pdf; 16,366 is the number of “OK” and “OKNTR” ballots from the file rejected & ok provisionals 04.pdf.

## References

- Särndal, Carl-Erik, Bengt Swensson, and Jan Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Wolter, Kirk M. 1985. *Introduction to Variance Estimation*. New York: Springer-Verlag.