THE HONORABLE JOHN E. BRIDGES

SUPERIOR COURT OF THE STATE OF WASHINGTON
FOR CHELAN COUNTY

Timothy Borders et al., Petitioners, v. King County et al., Respondents, and Washington State Democratic Central Committee, Intervenor-Respondent.	NO. 05-2-00027-3 DECLARATION OF DAVID J. BURMAN REGARDING WSDCC EXPERT HANDCOCK

DAVID J. BURMAN declares under penalty of perjury as follows:

1. I am one of the attorneys representing Intervenor-Respondent Washington State Democratic Central Committee ("WSDCC") in this litigation. I am competent to

Perkins Coie LLP
1201 Third Avenue, Suite 4800
Seattle, Washington 98101-3099
Phone: (206) 359-8000
Fax: (206) 359-9000

testify and make this declaration based on my personal knowledge and the files and records in this matter.

2. Pursuant to the Court's ruling of May 2, 2005, WSDCC hereby submits the report of WSDCC's expert Mark Handcock, Ph.D.

3. Due to the late disclosure of various supplemental reports by Petitioners' experts and supporting and related materials, and because the Court may reject certain categories of allegedly invalid ballots or challenges to individual allegedly invalid ballots, Respondent WSDCC reserves the right for its experts to apply the basic approaches set forth in their reports to the evidence submitted at trial.

SIGNED at Seattle, Washington, this 12th day of May, 2005

s/ David J. Burman
DAVID J. BURMAN

Perkins Coie LLP
1201 Third Avenue, Suite 4800
Seattle, Washington 98101-3099
Phone: (206) 359-8000
Fax: (206) 359-9000

Report on Statistical Aspects of the 2004 Washington Gubernatorial Election

Mark S. Handcock
Nannup Analytics

May 12, 2005

I was asked by legal counsel in this case to evaluate the reports of Dr. Anthony Gill (dated 15 April 2005, supplemented 21 April) and Dr. Jonathon Katz (dated 14 April 2005, supplemented 28 April and 4 May) on the impact of invalid ballots on the certified election results. My primary goal here is not to answer the question of who won the election. The primary scientific question I evaluate is whether the statistical evidence presented in the reports provides clear and convincing support for the contention that invalid votes changed the election result. As a standard for evaluation I have used what is generally accepted in the scientific community as the preponderance of evidence. This is a standard of proof falling somewhat below clear and convincing evidence.

A summary of my current findings about the reports is as follows:

- Neither the methodology that Dr. Gill or Dr Katz employs in their evaluations of the impact of invalid ballots on the certified election results is the standard used in the scholarly literature, nor has either achieved general acceptance. While their methodologies differ, both suffer from serious statistical defects that invalidate their main findings.

- The data used in both reports is neither a complete enumeration of invalid ballots (a "census"), nor a probability ("random") sample of ballots. As such, it is not possible to make a valid statistical inference from the data they have to the distribution of invalid votes in the election.

- Setting the data quality issue aside, the statistical methods used in these reports depend on assumptions that completely determine the outcome they obtain. As a result, their findings are not robust to small changes in these assumptions, and we can show that even slightly more reasonable assumptions (i.e., taking account of the sex of the invalid voter) produce an outcome that is the opposite of theirs. While it is not uncommon for statistical analyses to depend on the assumptions made, especially in the context of ecological inference, it is both uncommon and scientifically unacceptable to completely ignore the impact that such assumptions have on the outcome.

- The methods used by Gill provide no estimates of uncertainty, and therefore fall well below the standard used in scholarly statistical literature. Estimates of uncertainty, and principled statements of probability, are the defining feature of any statistical analysis.
- The methods used by Katz would, if used on the more complete data now available, and implemented with the alternative assumptions we examine, provide essentially the same results we present here.
- We can say with a high degree of certainty that the reports of Dr. Gill and Dr. Katz do not provide clear and convincing statistical evidence that the invalid votes changed the election result. In addition the reports do not show a preponderance of the statistical evidence supports this claim.

1 Qualifications

I am currently Professor of Statistics and Sociology at the University of Washington. I have been hired by legal counsel as a private individual for my relevant expertise, and do not represent the University. A complete copy of my curriculum vitae is provided as an appendix to this report.

I received my Bachelors of Science degree in Mathematics, with honours, from the University of Western Australia. I have a Doctor of Philosophy degree from the University of Chicago in the field of statistics. I have also done post-doctoral work at the IBM T. J. Watson Research Center. I became an Assistant Professor of Statistics in the Leonard J. Stern School of Business at New York University in September 1990 and was promoted to Associate Professor in 1994. In 1996 I accepted a position at the Pennsylvania State University with tenure. In 2000 I accepted a position as Professor of Statistics and Sociology at the University of Washington.

My graduate training was funded in part by a IBM Predoctoral Fellowship. My research has been funded by the National Science Foundation, the Rockefeller, Russell Sage and Ford Foundations, and the National Institutes of Health.

My work is largely motivated by questions in the social sciences. I have done extensive research developing statistical models for the analysis of social network data, spatial processes and longitudinal data arising in labor economics. I have co-authored three books on methodology and empirical research in the social sciences, including *Relative Distribution Methods in the Social Sciences*, and *Divergent Paths: Economic Mobility in the New American Labor Market*. The latter was awarded the Richard A. Lester Prize for the Outstanding Book in Labor Economics and Industrial Relations published in 2001 by Princeton University.

My research has been published in leading peer-reviewed statistical methodology journals such the *Journal of the American Statistical Association*, *Technometics*, *Sociological Methodology*, *Proceedings of the Royal Society*, and *Sociological Methods and Research*. I have published extensively on social science methodology and demography in leading peer-reviewed scientific journals such as *American Sociological Review*, the *Monthly Labor Review*, the *American Journal of Sociology*, *Demography*, and the *Journal of Labor Economics*. In addition I have published on statistical methodology in leading peer-reviewed interdisciplinary scientific journals such as *Nature*, *Ecology*, and *Theoretical Population Biology*.

A particular interest of my work has been statistical methodology for the combination of sample survey and population-level information. I am currently funded by the National Institute of Child Health and Human Development to develop statistical methodology in this area with application to demography. I have served as the Director of the Statistics Core for the population research centers at both Penn State University and University of Washington. These Cores provides statistical guidance for demographic research including study planning and survey design and analysis.

I am a core member of the Center for Statistics and Social Sciences at the University of Washington. Part of the mission of the Center is to galvanize collaborative research between social scientists and statisticians.

I am an expert in spatial statistical methods, and have published on spatial distributions and inference for spatial processes. The treatment of the spatial distribution of invalid ballots is important to the determination of their impact on the outcome of the election.

2 Evaluation of the Data Underlying the Katz and Gill Reports

Statistics as a science is concerned with the evaluation of the weight of evidence for and against hypotheses. The key components of a statistical analysis are estimation and inference. Both depend on the nature of the evidence that will be analyzed, so the initial phase of any evaluation is to understand the quality of this evidence. Where do the data come from? Is it representative of the population of interest? Is there information in the data that can be used to assess the hypotheses? Based on the answers to these questions, statistical methodologies are used to estimate quantities that explicitly test the hypotheses. The tests, and the inferences we make from them, typically require additional assumptions. If the assumptions are met, the methodologies are robust, unbiased and reliable. Finally, the weight of the evidence is assessed. Is there enough data to provide clear and convincing evidence? What is the uncertainty in our estimates? How sensitive is the analysis to the assumptions, and how robust are the findings? Is there any evidence in the data to support the assumptions we have made; is there any evidence that can be drawn from external information?

To evaluate the scientific validity of the reports we review each of these steps.

In this section we consider the nature of the data set used and how it relates to the findings claimed in both reports. The approaches used in both reports are based on many assumptions. We focus on the two key assumptions they each make about the relationship of their data to the votes cast in the election: the fraction of all invalid votes that they have, and who these invalid voters voted for. Both assumptions need to be correct for their findings to be valid. Both assumptions have a large impact on the inferences they can draw from their findings. As we shall see, neither has much evidential support or even face validity.

2.1 Assumption I: Their data set is a known fraction of the invalid voters

This assumption states that the list of invalid ballots in their data set is either: (1) a complete enumeration of the all invalid ballots cast in the election, or (2) a sample of the invalid ballots, where the probability of each ballot entering their sample is known.

The first approach would claim that the set of invalid ballots in their dataset is a true *census* of the invalid votes cast. This means that they have every invalid voter and have not included any valid voters as invalid voters. If this assumption was correct, then there would be no question of "design based" statistical inference from our sample to the population of invalid voters: this is the population of invalid voters. Assuming for a moment that we knew how these invalid voters had actually voted, then we would be able to say with certainty whether these votes had changed the election outcome. The findings obtained from our data would be the findings of interest, there would be no need to evaluate whether an inference can be made from these data to the set of all invalid votes (and by extension, to the election outcome).

The second approach would claim that, even if we do not have a census of invalid voters, we have a probability or "random" sample of these voters. In this case, we can use statistical methods to make this sample representative of the population of invalid voters, estimate the votes they represent, and make the inference from this sample to the set of all invalid votes, and thus to the election outcome. For example, if every invalid voter is equally likely to show up in our sample, and we know the fraction of all invalid voters that our sample represents, our data would comprise a "simple random sample" of the invalid votes. If we assume again that we know how these invalid voters actually voted, it would be a simple matter to estimate the impact these votes had on the election outcome. It is not necessary that each invalid voter have the same "inclusion probability" in our sample, it is only necessary that we know what that inclusion probability is in order to adjust for it.

If we can claim neither (1) nor (2) then we have a statistical problem. The statistically accepted procedures for calculating an unbiased estimate from a sample requires the probability of inclusion of the sample elements to be known (Kalton 1983).

The original data for the Gill and Katz reports was provided by Polidata. The number of cases in this data set has changed over time, and the counsel for Gregoire has identified an additional 743 probable invalid felon voters, primarily in different precincts than the Polidata data, nearly doubling the number of cases. Their list also identifies 1898 provisional ballots counted before verification outside of King County, and 780 absentee and provisional ballots from King County that are identified as not being counted due to error by election officials. The legal status of the invalid voters and the provisional ballots from both sides is still to be determined. It is therefore clear that Gill and Katz were not working with a census of the invalid voters.

This means they must be working with a sample, the only question is what kind of sample.

The data set used by Gill and Katz does not appear to be based on a scientific sampling design that produces known inclusion probabilities. If it had such a design, the data set would have included a weighting variable for calculating valid population-based estimates. There is no such variable in their data set. In the absence of a scientific sample design, these data would be referred to as a *convenience sample* in the statistical sampling literature (Kalton 1983). Hence we shall refer to it as the *Rossi convenience sample.* This means that we have no scientific way to determine how the sample relates to the true census of invalid voters. The use of a convenience sample is widely regarded as inappropriate when the goal is to make valid inferences to a larger population, and it is not acceptable in the scientific community as a basis for clear and convincing evidence.

For this reason, we shall also refer to the additional invalid and provisional ballots identified by the counsel for Gregoire as the Gregoire convenience sample. We refer to the pooled data sets as the *combined sample.* For the purposes of the numerical analysis in this report we must make a choice about the specific data to use. As our primary intent is to evaluate the statistical evidence presented in the Gill and Katz reports we will use their complete list (Table 1 of the Supplemental Report of Katz, 28 April 2005). This includes 943 invalid felon voters, 56 other invalid voters and 174 provisional ballots. The combined sample includes 743 invalid felon votes identified by the counsel for Gregoire, but does not include any of the many provisional or absentee ballots they have also identified. Thus the combined sample includes all those used by Katz and only the invalid felons identified by the counsel for Gregoire. As the legal status of the invalid voters and the provisional ballots from both sides is determined, we expect to submit supplemental reports using the updated data.

We emphasize that the combined sample is not a census, nor does it have a set of known inclusion probabilities. It is simply the most complete data set available at this time.

Katz and Gill do not explicitly claim that the Rossi convenience sample is a census of the invalid voters. In fact, they explicitly state that they only analyzed the data provided to them, referring all questions about the data to Polidata. What they neglect to say is that no valid statistical inference can be drawn from these data without making either assumption (1) or (2) above.

This is not acceptable statistical practice.

2.2 Assumption II: The invalid and valid votes are exchangeable within precincts

Suppose it were possible to obtain a census of the invalid voters – we still would not know how those invalid voters actually voted. The second assumption made by both Gill (15 April, Section B.2) and Katz (14 April, Section 4.1) is that the ballots of the invalid voters are "exchangeable" with those of the valid voters within the precincts they voted. That is, we assume the invalid voters voted just like anyone else in their precinct. This is also known as the "ecological inference". The reports give a very vague discussion of the nature and basis for this crucial assumption.

The notion of exchangability has a long history in the field of statistics (de Finetti 1930). The implicit specific assumption made by Katz and Gill is that the votes on all ballots cast are finitely *partially exchangeable* with respect to precinct (Bernardo and Smith 1994).

The assumption of exchangability should only be made when there is strong contextual evidence supporting it. In this case, that the votes on the invalid ballots are "homogeneous," "indistinguishable," or "just like" the votes on the valid ballots. Usually this would imply that we think the factors that influence voter choice are similar for both, and there are not other characteristics of the voters that would enable us to distinguish between the votes.

This is *prima facie* not true in this situation. Many of the characteristics of the invalid voters are clearly distinguishable from the valid voters. The invalid ballots are cast by in-state or multi-state dual voters, in the name of deceased voters, and felons who did not have the right to vote. Katz and Gill do not provide substantial contextual evidence that the choices of these individuals are indistinguishable from those of valid voters. Indeed they provide very little evidence at all - they claim it without support. Making such a claim, without strong contextual evidence, is neither standard, nor an acceptable statistical argument to make.

Note that the issue is not whether we can distinguish the votes now that they have been cast, but concerns the actual votes cast by invalid voters. These two issues should not be confused: a ballot randomly drawn from those cast in the precinct does follow the probability distribution of the precinct as a whole and is exchangeable. However we do not have random drawn ballots here. The issue is if the actual votes of the invalid voters are just like those of the valid voters. The scientific question relates to the actual number of valid votes cast for the candidates, not if we can distinguish the invalid from valid votes once they have been cast.

The standard and acceptable statistical approach to this issue is to use *missing data* methodology (Little and Rubin 2002). The missing (i.e., unknown) votes in this case are those on the invalid ballots. If we knew the votes on the invalid ballots we could subtract them from the totals and resolve the issue. So how do we estimate these missing votes?

A core question in missing data methodology is whether the missing data mechanism is "ignorable" - that is, are the cases with missing data "like" the cases with complete data. In the weaker form of this assumption, the data are assumed to be "missing at random" (MAR), once you take into account other measured characteristics. In this context, it would mean that, even though invalid voters might be more likely to be male, white, older, and wealthy, they vote just like valid voters who are male, white, older, and wealthy.

Katz and Gill make the much stronger assumption that the invalid votes occur *completely at random* (MCAR) within the precinct (Little and Rubin 2002). That is, the invalid votes are not affected by the characteristics of the voter within the precinct. This is a very strong assumption and is not an acceptable statistical argument to make unless there is strong evidence to support it.

This assumption is *prima facie* incorrect in this situation. Invalid voters differ in many key characteristics from the valid voters. The fact that they are invalid suggests that their vote may differ from the precinct propensity to vote for Gregoire, Rossi, Bennett or other. In particular we have evidence that they are more likely to be male than the typical voter in the precinct. So at a minimum, we should examine whether the results we estimate based on the very strong MCAR assumption (that there are no sex differences in voting patterns), are robust to the weaker, but more realistic MAR assumption (that voting differs by sex, but it otherwise similar).

More generally, it must be recognized that any results predicted by these methods are going to be strongly influenced by the assumptions we make at this stage about how the invalid voters voted. So the assumptions must be carefully examined, their influence on the results tested through sensitivity analyses, and weaker assumptions should be preferred to stronger in the absence of clear convincing evidence to the contrary.

If the exchangeability assumption is false, what can we say about how invalid voters voted? Gill (Section E) and Katz (April 14, Section 4.2) claim that felons are more likely to vote for Gregoire than they would under the exchangeability assumption. However they do not use this information directly in their analysis. The only evidence they cite for this claim is the sociological study Uggen and Manza (2002). This is a misrepresentation of Uggen and Manza (2002) and the evidence. Uggen and Manza (2002) does not have data on the votes of felons but assume that felons vote like non-felons with similar demographics. We return to this issue in Section 4.3.

3 Evaluation of the Technical Aspects of the Reports

As the Rossi convenience sample is incomplete, and its relationship to the census of invalid voters is unknown, it is very difficult to use it as the evidential foundation of a statistical argument that illegal votes changed the election result. To do so requires the specification of very strong and untestable assumptions of the kind that have not achieved general acceptance in scholarly literature. If these assumptions are made then the relevance of the analysis to the primary question will be greatly reduced: the results will simply reflect the assumptions, not the data.

In the remainder of this section we undertake a counterfactual analysis where we assume that the combined convenience sample is a census of the invalid voters. We do this to demonstrate the sensitivity of their results to their assumptions, even if they had been applied to a census.

As is described in Section 4.1 of Katz's report, in this situation we are attempting to make an *ecological inference* based on aggregate data. Katz gives a clear overview of the statistical issues in Section 2.1 of his "Report on the Georgia Congressional and Legislative Redistricting". The group of voters in that case are minorities, while in our case it is invalid voters. In that case the minority make-up of the precincts are assumed known. As we note in Section 2, in our case the invalid/valid make-up of the precincts is not know as we do not have a census.

In his report Katz describes the work of Goodman (1959) as the standard technique for ecological inference. However there have been many well known and well accepted developments in this field since then. For the redistricting case Katz notes and advocates the method of bounds and the EI approach of King (1997) as accepted methods. More recently Wakefield (2004) developed a binomial convolution model that respects these bounds, and King *et. al* (2004) reviews other recent advances in these methods. These newer methods of ecological inference are very applicable to the primary question addressed here: Is there sufficient information in the data available to infer that that illegal votes changed the election result? If the method of bounds were applied it would indicate that the data do not bound the distribution of votes by invalid voters at all well. If the more sophisticated parametric models of Wakefield (2004) were applied they would almost certainly show that any statement about the patterns of invalid votes depends critically on the assumptions of the model as the data is not informative. This addresses the primary question by assessing the informational content of the data, and systematically demonstrating that the evidence provided by the data is weak.

It is now standard in situations of ecological inference to undertake a sensitivity analysis of the model to the assumptions (see Section 10 of Wakefield 2004, and the discussion following it). Neither Katz nor Gill undertake a sensitivity analysis. We provide one in Section 4.

3.1 Statistical Analysis Based on the Exact Distribution

Both Gill and Katz use a number of approximations in their methodology that are not necessary in this case. To see this we review the standard statistical approach to this problem. Throughout this section we assume that we have a census of the invalid voters and that the invalid and valid votes are exchangeable. As noted in Section 2, there is strong evidence that these assumptions are incorrect, and neither has been given empirical support by Gill or Katz. But we make the assumption here in order to focus on the methodological issues.

Consider first the statewide set of ballots. Let G, R, and C be the valid votes for Gregoire, Rossi and the total number of valid ballots cast, respectively. Let $O = C - G - R$ be the number of valid ballots cast for neither Gregoire or Rossi (e.g., for Bennett or over/under votes). Suppose there are m invalid ballots cast.

The assumption of exchangability of the ballots is that the set invalid ballots are equally likely to be any of the possible sets of size m from the total set of ballots cast. Let $S = \{S_1, \ldots, S_m\}$ be the indices of the invalid ballots, and $\boldsymbol{S}$ be the set of possible sets of indices. Under the assumption of exchangability S is a random variable equally likely to be each of the $\binom{C+m}{m}$ choices in $\boldsymbol{S}$.

Let $G(s)$ be the number of valid votes for Gregoire if the index of the invalid votes is s. Similarly let $R(s)$ and $O(s)$ be the number of valid votes for Rossi and Other, respectively. For each set of invalid ballots we can determine the vote totals of the valid ballots and hence determine Gregoire's margin of victory and the victor.

The primary quantity of interest is if the result of the election would change, that is, is $R > G$. Under the assumption of exchangability, each of these electoral outcomes is equally likely to be the actual outcome. Based on this, the probability that Rossi exceeds Gregoire is:

$$P\Big(R(S) > G(S)\Big) = \frac{1}{|\boldsymbol{S}|} \sum_{s \in \boldsymbol{S}} I\Big(R(s) > G(s)\Big) \tag{1}$$

where $I(A)$ is 1 if the statement A is true and 0 otherwise. Under the assumption of this section this can be computed using a standard Monte Carlo algorithm[1].

Consider now the analysis at the precinct level. The assumption of exchangeability means that the set of invalid ballots for that precinct is equally likely to be any of the possible sets from the total set of ballots cast in that precinct. The distribution of S now has components for each precinct. Again assuming the number of invalids cast and the respective totals of ballots cast at the precinct level are known, this can be computed based on formula (1) using the precinct level version of $\boldsymbol{S}$.

To illustrate this approach, we apply this methodology to the combined sample of invalid voters (defined in Section 2.1). This is not a census as the unknown invalids and provisional ballots identified by the Gregoire counsel are excluded. However it is more complete than either convenience sample alone. Figure 1 is the plot of the distribution the Gregoire margin of victory (i.e., $G(S) - R(S)$). The distribution is over the count of the number of votes. The mean of the distribution is a 67 vote victory for Gregoire, but the more important information is the other relevant probabilities that can be read from the plot. The primary probability of interest is the proportion of possible scenarios that lead to a Rossi victory. This is 4.6%. In words, under the scenario where the combined sample is a census of the invalid voters and their votes are exchangeable with the valid votes at the precinct level, the chance that the elimination of the invalid votes will change the election result is less than one in twenty.

As this is based on assumptions that we know are not correct we do not claim that these calculations are correct. However, under the assumptions as stated by Katz and Gill this is a standard and accepted computation of the probability.

In contrast, in his April 28 report Katz uses the Binomial approximation to this distribution, followed by the Normal approximation to the distribution of the invalid vote counts to construct an interval of value he refers to as a "95% confidence interval for the estimate." Note that this is not a standard usage as it does not express uncertainty resulting from sam-

[1]Explicitly, we randomly draw $M = 100,000$ elements from $\boldsymbol{S}$ and use the sample mean $\frac{1}{M}\sum_{i=1}^{M} I(R(s_j) > G(s_j))$ as a numerical estimate. The Monte Carlo error in the estimates reported here is less than 0.002 and can be made arbitrarily small by appropriate choice of M.

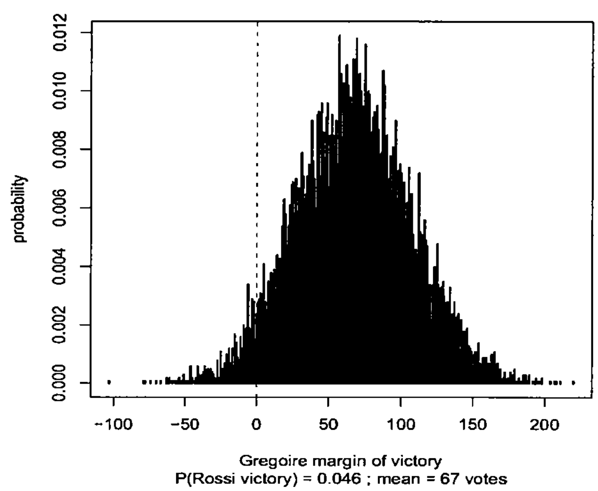

Figure 1: Distribution of the Gregoire margin of victory over all possible sets of invalid votes in the combined sample. This is the scenario where the combined sample is a census of the invalid voters and they are exchangeable with the valid votes at the precinct level. The dashed vertical line corresponds to a tie. The proportion of possible invalid votes that lead to a Rossi valid is 4.6%.

pling the invalid votes. It is an approximation to a probability interval completely specified by the assumptions. Under the exact distribution of Figure 1, the analogue is the smallest interval that has probability 95% of containing the actual Gregoire margin of victory. It is from -9 to 146 votes here. Note however that Katz's interval is an indirect measure of the effect of the invalid votes. A more direct measure is the probability of a changed result – 4.6% in this case.

In his Second Supplementary report of May 2, Katz appears to have updated his approximation to the multinomial incorporating the votes that were not for Gregoire or Rossi. Our analysis does this directly. When we apply his methodology to the combined selection we find it provides essentially the same results as for the exact method we apply here.

The details of the methodology in Gill's report are based on a so called "tipping point" analysis. It is similar in spirit to the method of Katz but much less statistically informed. His baseline tipping point analysis uses the deterministic rule based on a point estimate of the mean, requiring only that there are enough votes that this could happen. If we use his approach on the combined sample, his probabilistic tipping point rule essentially asks whether the mean of the distribution in Figure 1 is negative (the "tipping point"). In this case it is positive. The deficiencies of the probabilistic tipping point analysis are clear: It does not take into account the probability of being negative, and hence does not assess the weight of evidence that the result of the election will be changed. To see why this is important, consider the situation where the mean of the distribution is -1, but almost half of the values are positive. Then the probability the result of the election will be changed will be about 50% even though the probabilistic tipping point analysis claims there is strong evidence.

I do not know of a usage of this probabilistic tipping point methodology in the scientific literature that meets current standards. Even when the strong assumptions it requires are met, it is technically inferior to the exact method given in this section.

4 Evaluating the Robustness and Reliability of their Results

It is usual in a statistical analysis of evidence to assess the sensitivity of the results to the assumptions made. Katz and Gill do not include such sensitivity analyses in their reports. Gill does not even include measures of statistical variation.

In this section we conduct a number of sensitivity analyses on the assumptions used in the previous sections. We do so by considering alternative assumptions and their effect on the assessment of the evidence. The analyses presented here are by their nature exploratory and not definitive. We consider three scenarios: What is the impact of adjusting the demographic characteristics, how similar are the convenience samples to the census of invalid felon voters, and what is known about how invalid ex-felon voters vote.

4.1 Adjusting for the Sex of the Invalid Voters

.

In this subsection we consider the impact of the demographic characteristics of the invalid voters on the results. Our objective is to see how alternative assumptions influence the results. These assumptions are at least as reasonable as those made in the reports, and are arguable more so. However, we do not argue that this is the only, or even best, way to statistically model the process.

I have been provided by counsel with the sex of each of the invalid voters in the Rossi selection and the complementary selection. A high portion of the invalid voters are male (75%). It might be expected that male invalid voters are more similar to male voters within their precinct than female voters. This is a refinement of the exchangeability assumption to take into account the heterogeneity of the invalid voters.

There is also a differential reported in the National Election Poll (`www.exit-poll.net`) between the proportion of males at the statewide level who voted for Gregoire (53%) and

	Vote			Total
	Gregoire	Rossi	Other	Vote
Male	45%	53%	2%	45%
Female	53%	46%	1%	55%

Table 1: National Election Poll results for the Washington Gubernatorial election. The sample size is 2,152. For their methodology see `http://www.exit-poll.net`.

females (45%). We report these values in Table 1.

As this is an exit poll it is subject to the usual errors as an estimate of the actual vote. In addition we do not have the sex differential at the precinct level. However, we can check the sensitivity of the results to the sex differential by applying hypothetical values at the precinct level and evaluating their effect.

Suppose that the ratio of the probability that a male votes for Gregoire, Rossi, or neither within a precinct to a female is the ratio given in the National Election Poll. This is an ecological assumption. We can then determine the number of male and female voters for each candidate by arithmetic in such a way that the vote totals are preserved. In essence we have divided each precinct into a precinct for males and a precinct for females. I have also been provided with the proportions of male and female voters in each precinct. We can then apply the exchangability assumption of Section 2.2 to this case, and calculate the effect on the vote totals and the probability of a changed election result. The set of possible valid totals is given in Figure 2. We see that adjusting for the sex of the invalid voters has a large effect. The probability of changing the election result is now close to zero. In fact, Gregoire is likely to increase her winning margin by an average of 31 votes to an average of 160 votes.

While this scenario should not be regarded uncritically, it is at least as compelling as the model that ignores the sex of the invalid voters. It also indicates that the likelihood of a changed election result is very sensitive to small variations in what we assume about the invalid voters. Although we have some information on the sex differential, we do not know about the other demographic characteristics that effect voting. This is because we do not know the characteristics of the invalid voters, and more importantly do not know how invalid voters vote given those characteristics.

4.2 Evidence for where the Ex-felon Invalid Voters are

We now return to an issue addressed in Section 2.1. The combined selection is not a census of the invalid voters. How similar is it to the census of invalid voters? Without the census this is difficult to answer directly. In this section we address this question indirectly by looking at a proxy for the number of invalid felon voters. The ex-felon databases that both

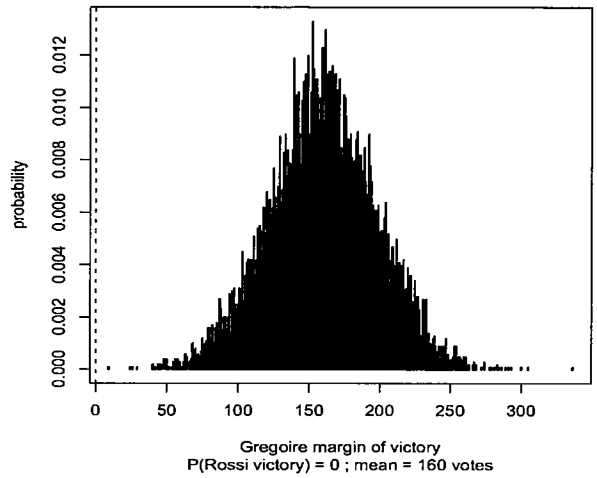

Figure 2: Distribution of the Gregoire margin of victory over all possible sets of invalid votes. This is the scenario where the combined selection is a census of the invalid voters and they are exchangeable with the valid votes at the precinct level. The dashed vertical line corresponds to a tie. The proportion of possible invalid votes that lead to a Rossi valid is 4.6%.

sides are using to compare to the 2004 voter databases does not have address information; the address information for the ex-felons who apparently voted is added from a registered voter database. We consider data on the numbers of felons under supervision by the State of Washington Department of Corrections for Fiscal Year 2005. Parolees typically receive 36 months of community supervision and so this measure focuses on the immediate post-prison years. The information is available at the county level (Department of Corrections 2005).

We start by considering the number of felons under supervision per cast ballot in each county. A key variable driving the assessment of the impact of invalid ballots on the certified election results is the percent of Gregoire vote. If more invalid felons per capita occur in counties who strongly supported Gregoire then the removal of the invalid ballots will help Rossi. To place these on a interpretable scale we divide by the value for King County, and refer to it as the *relative density of felons*.

Figure 3 plots the relative density of felons under supervision against the percent Gregoire vote for each of the counties in Washington. For example, Lewis County had a Gregoire vote of 32% and per-capita felon 2.52 times that of King County.

The plot does not indicate a propensity for felons to be supervised in counties that support Gregoire more. In fact the smooth curve superimposed on the scatterplot suggests that the representation of felons decreases slightly as the support for Gregoire increases.

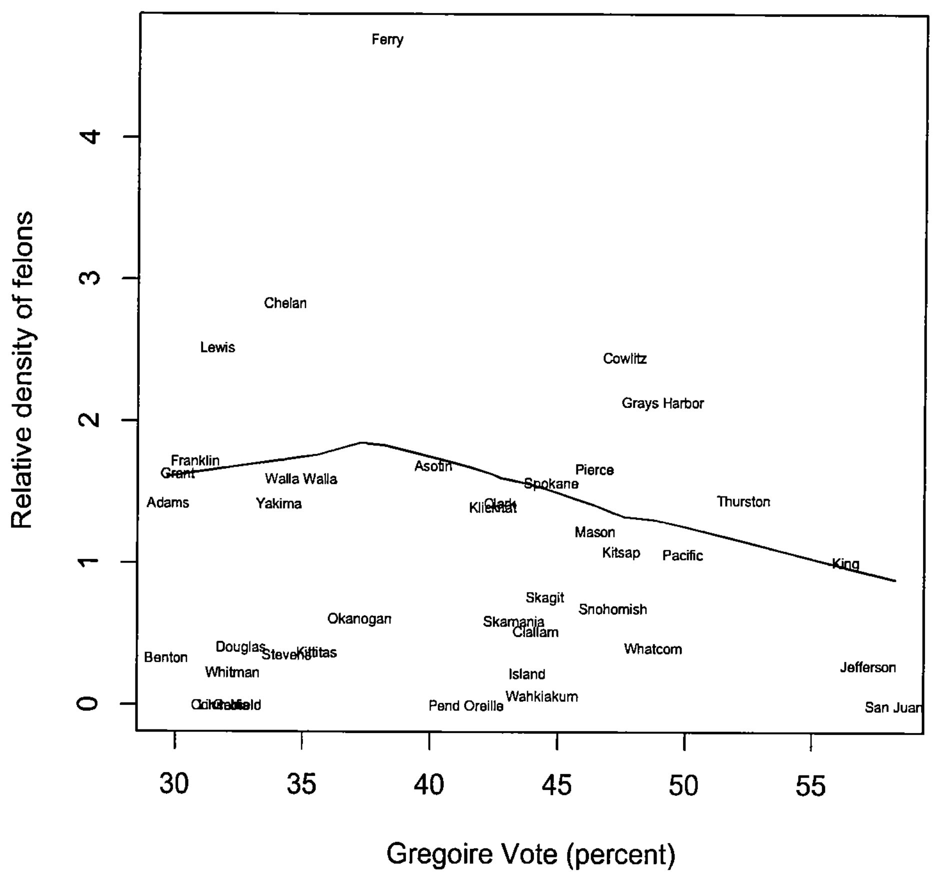

Figure 3: Distribution of the clients under supervision by the Department of Corrections by percent of vote for Gregoire for counties in Washington. As the percent vote for Gregoire increases the per-capita felons trends down modestly. King County ranks in the middle in terms of felon representation. A smooth curve is superimposed on the scatter plot.

Next we consider the propensity for the Rossi selection to report felons within a county. For each county we consider the ratio of the number of felons reported in the Rossi selection for that county divided by the number of clients under supervision in the county. To place these on a interpretable scale we again divide by the value for King County, and refer to it as the *propensity to report invalid felons.* To the extent that the number of invalid felon voters is proportional to the number of clients under supervision we may expect this ratio to be approximately one. If a county has a relatively high ratio it means that the Rossi selection has more invalid felons than expected under proportionality.

Figure 4 plots the propensity for the Rossi selection to report felons against the percent Gregoire vote for each of the counties in Washington. For example, Benton county had a Gregoire vote of 30% and a propensity for reporting felons 30% of the King County level.

The steep increase in propensity as the support for Gregoire increases above 45% is apparent. This is largely driven by the dramatic over-representation of felons in King County relative to the other counties.

We repeat this analysis for the combined selection. Figure 5 plots the propensity for the combined selection to report felons against the percent Gregoire vote for each of the counties in Washington. We see that the combined selection has a polarization of reporting with both low and high support counties being overreported relative to the medium counties. This suggests, but does not prove, that there are more invalid felon voters to be identified in the mid-range counties. The inclusion of these invalid felon voters, if they exist, would bring the list closer to the census of invalid voters. It also emphasizes that the combined sample, and certainly the Rossi convenience sample alone, is not a census, nor representative of all invalid voters.

As a final stage of this sensitivity analysis of the invalid ex-felon voters we constructed a synthetic census of ex-felons by proportionally adjusting the numbers of invalid felons voters to be same as the per-capita felons under supervision (Figure 3). The absolute level of invalid felon voters is adjusted up to that of King County as this county appears to be the one most exhaustively investigated for invalid felon voters. Based on this synthetic census we then can apply the exact methodology in Section 3.1 to assess the probability of that the invalid votes changed the election result. When we do this we find the probability is small, especially if we apply the demographic adjustment for the sex of the invalid voters given in Section 4.1. This is an exploratory analysis premised on a possible census of invalid voters.

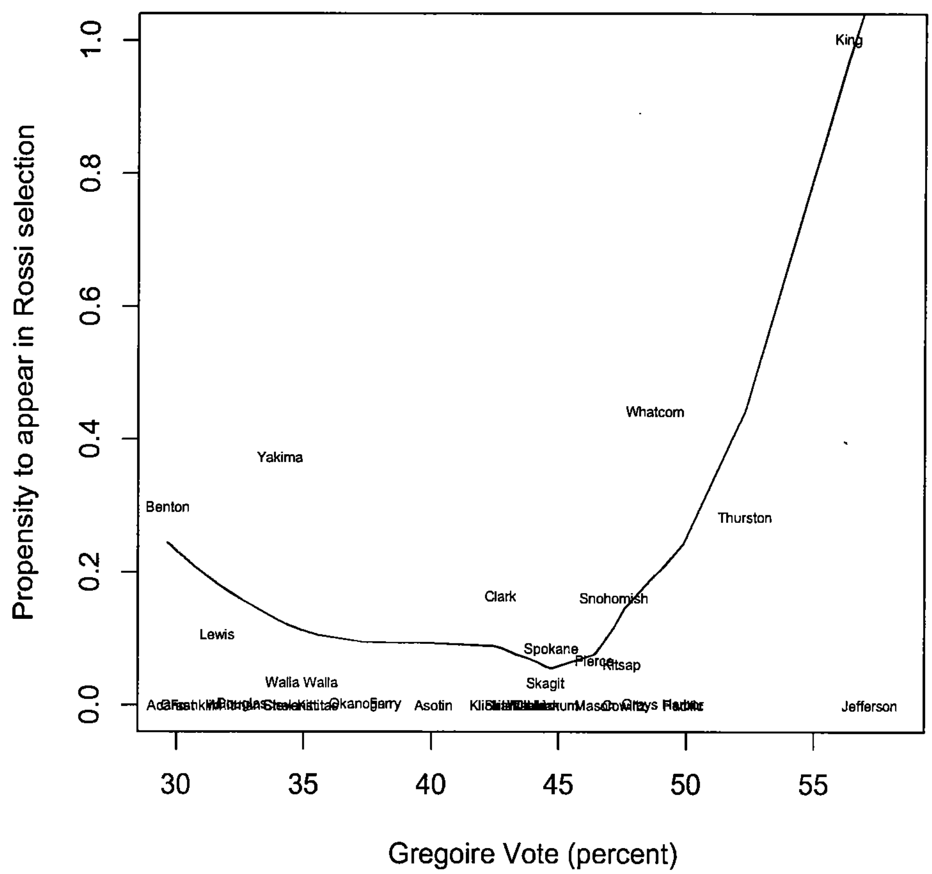

Figure 4: Distribution of the propensity for felons to be reported by the Rossi selection by the percent of vote for Gregoire. All values are relative to the level for King County. There is a steep rises in reporting when the percent vote for Gregoire increases above 45%. A smooth curve is superimposed on the scatter plot.

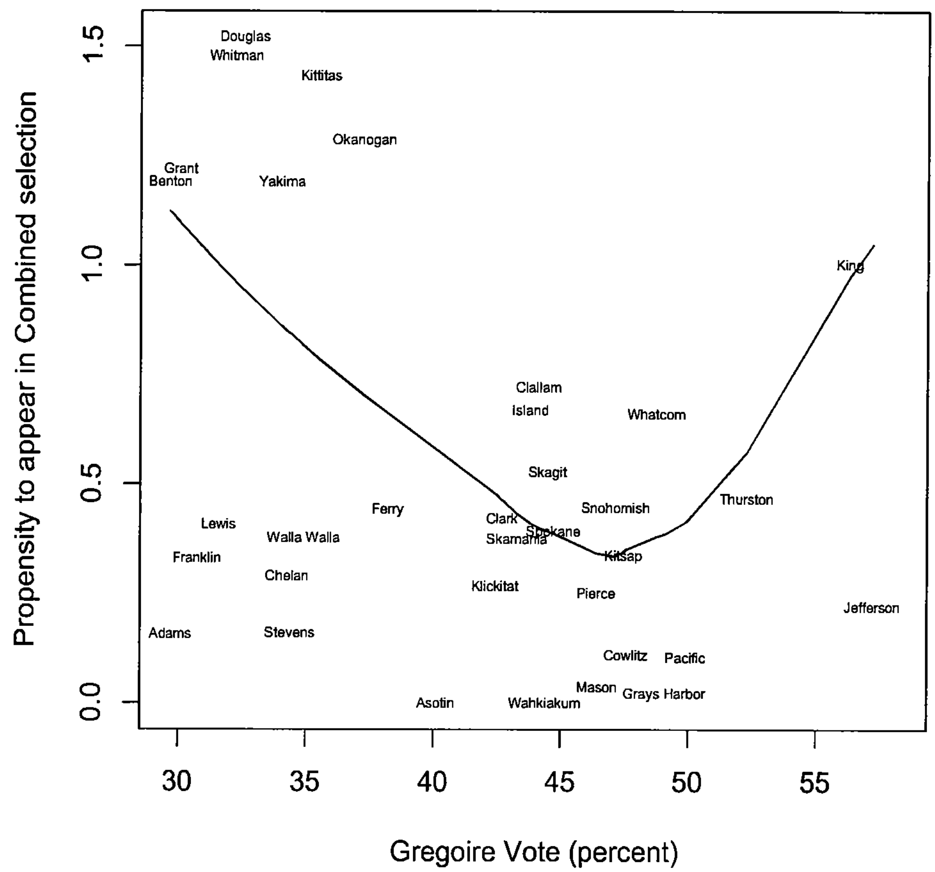

Figure 5: Distribution of the propensity for felons to be reported in the combined selection by the percent of vote for Gregoire. All values are relative to the level for King County. The complementary selection has polarized representation of low and high Gregoire support counties. The combined effect is to greatly underrepresent counties in the middle range of support. A smooth curve is superimposed on the scatter plot.

4.3 Evaluating the Evidence for how Invalid Ex-Felons Voters Vote

We now return to an issue briefly discussed in Section 2.2. Throughout this report we have assumed, as do Katz and Gill, that invalid voters vote like valid voters in the precincts where they vote. In this section we review the evidence for how felons vote. The only evidence Katz and Gill introduce is the sociological study Uggen and Manza (2002). What evidence is there in Uggen and Manza (2002) for the way invalid felon voters in the 2004 Washington gubernatorial election voted?

Some important issues are:

- Uggen and Manza (2002) does not report data on how felons voted.

  They are using the "National Election Study" – a random sample of about 2500 Americans and predicting how the respondents in this sample vote, as a function of demographic characteristics. To quote from Uggen and Manza (2002), p 784:

  > "To analyze the expected turnout and vote choice of disenfranchised felons, we do not have any survey data that asks disenfranchised felons how they would have voted."

  Also from a footnote on p 784, note that:

  > "... the National Election Study ... typically overestimates turnout by 18 to 25 percent."

  They do have a much smaller sample of youth (about 300) from St. Paul who have been either arrested or incarcerated. In this sample, none of the demographic characteristics consistently predicted party preference, and the only criminal characteristic that predicted party preference was as strongly related to independent preference at the state level (Ventura) as to democratic at the national (Clinton). The relevant data are in Table 3, page 791, last 2 columns.

- Uggen and Manza (2002) do not report how well their model predicts voting intention for this sample.

  This is a major oversight in the publication of this paper. It is standard in this kind of analysis to report a statistical measure for the "goodness-of-fit" of a model. Uggen and Manza (2002) do not do this. For logistic regression, the standard measures are either the reduction in the log-likelihood, or a classification table.

  The relevant table is in Appendix 3, page 799. Coefficients for each model (for each year) are shown, but there are no fit statistics.