Notes about states or CDs from the second pass:

This pass follows the release of a full nationwide dataset which included data by CD and by County within CD (CC) for both the current (2010, C100) plan and the new (2012, C199) plan.

The main focus of the first pass was to locate, convert, and process a disaggregated dataset for each state so that a PVI could be calculated for the current plan and if comparable to the original PVI from the post-election PRCD projects, to do the same for the new 2012 plans. The focus for the first pass was on the overall behavior and not so much on the raw votes because the first was needed in a timely fashion but the second was not needed until after the election when they would be published in the Almanac.

The main focus of the second pass was to review the results of the first pass to address instances in which the difference between the original PVI and the recalculated PVI for the current plan was greater than 0.5 raw score or 1 whole point. Most of the CDs with a 1 point difference in the PVI were indeed based upon a raw difference of less than one full point but ones that moved past the 0.5 rounding point. (For example, IN—4 (Rokita—R) was 14.3 (R+14) originally but 14.5 (R+15) recalculated). (There were two other types of CDs to review: a) those with more than a 1 point change in the PVI (only 2 CDs in AL and NY); and b) those with a 1 point change in the PVI that did not flip solely due to rounding (i.e., generally those with a change in raw vote of more then 0.5).

Considering the differences in the datasets used (for the post-election PRCD projects versus the pre-2012 PRCD project) and the allocation and matching issues, some minor differences are to be expected. Nevertheless, these needed to be reviewed both to answer the question “Is the underlying disaggregated dataset up to the task of providing political indicators for the new plan?” as well as reviewing the raw vote totals.

To the extent it was practical to do so, absentees or other centrally-reported votes were allocated in the disaggregated datasets, either by the vendors that produced them or by Polidata after the fact. Allocation was, however, only possible after the fact if a) a PDset/SRset for the state/year was available; and b) the designation of absentee votes in the PDset approximated the number of votes missing from the disaggregated set. In other words, I could not allocate absentees into the DISAGGset from a PDset if the DISAGGset was missing 250,000 votes but there were only 50,000 votes designated as being absentees in the PDset.

There were two basic options for the allocation of the absentees: a) by county; and b) by CD within county (CC). In most cases the allocation was done by county because this was the most expeditious for the first pass. For the second pass the allocation by district was investigated but this was full of problems due to both inconsistencies in the coding of the PDsets and the difference in the way local election authorities (LEAs) reported the results. (To date updates were made, and retained by virtue of their improvement, for CA and VA based upon a refined allocation of centrally-reported votes.)

Another reason for the differences that exist in the pre-2011 PRCD numbers and the post-election PRCD projects is the matching aspect of the DISAGGsets. As this was done by the vendors who built the sets, it

will require investigation into the matching information to determine whether the differences can be addressed. This will take some considerable amount of time.

The important point to understand here is that just having a precinct-level election dataset with all votes accounted for is not enough to solve these types of issues. The votes need to be matched to a 2010 census block to which estimates of votes are allocated based upon a population factor. Each block thus carries forward an estimate of votes so that when it is assigned to a district in a new plan, the estimates are aggregated with all other blocks in the district. Matching for this purpose indicates the assignment of each census block to a precinct for each election cycle. By the very nature of undertaking this huge task, there will be some differences in vote totals that are based upon the estimates block by block.

The issue of the raw vote differential was not much of a concern for the first pass because, to the extent that all votes were accounted for, the differential in raw votes should not be very great. The focus then was on comparing the PVI values and though some differences were noted, they were left for the second pass. A review of this for the second pass located instances in which there was a differential and in several of these, offsetting differentials, i.e., District A was under by about 10% and District B was over by about 10% in total votes. For example, in NY—2 (Israel—D), NY—3 (King—R), and NY—4 (McCarthy—D), the comparison of the votes for the 2008 election indicates an issue with respect to the raw votes: Israel is under 10%, King is under 19%, and McCarthy is over 28% when comparing the raw total vote count to the original post-election numbers. The post-election PRCD numbers make sense for these CDs so the issue appears to be either a matching issue or an allocation issue. Unfortunately, I do not yet have a PD/SRset for either year but the DISAGGset does not appear to be missing enough votes to fix this and therefore it would appear to be a matching issue.

There are a few states for which there will be some differential in the raw votes due to the inability to allocate absentees or centrally-reported votes. These include at least the following: CA (all CDs with a portion of Los Angeles County and a handful of others); CT, MA, NY, OR, and UT.

In addition, there are specific CDs which have an issue with respect to the raw vote totals. The differences may or may not affect the recalculated PVI due to the status of the differential and whether they are due to matching or allocation issues. For example, the differential between the parties for absentees was not as great for 2004 and it was for 2008. The following districts have raw vote differences that need more investigation.

AL—6 and 7; check Jefferson;

FL—3 and 4 have offsetting issues for 2008 (+3.5 and -3.6), check Duval;

FL—17 and 18 have offsetting issues for 2004 (+8.6 and -6.9), check Miami;

GA—4 and 5 have offsetting issues for 2008 (-5.5 and +5.4), check Fulton, DeKalb;

GA—6 and 11 have offsetting issues for 2008 (+2.5 and -3.0), check Cobb;

GA—13 is short for 2008 (-9%), check Fulton;

MA—1 and 2 are short for 2008 (-9.4 and -11.6), check Hampden;

MA—8 and 9 have offsetting issues for 2004 (+7.1 and -7.3), check Suffolk;

NY—2, 3 and 4 have offsetting issues for 2008 (-9.8, -19.2, +27.7), check Nassau;

NY—26, 27 and 28 have offsetting issues for 2008 (+2.8, +5.5, -15.5), check Erie, Monroe.

My suggestion is that the raw vote totals for these districts, if listed at all, accompany this disclaimer:
"The raw vote totals for this district are still under review.