THE IMPACT OF SAMPLE ATTRITION AND RETROSPECTIVE RECALL BIAS ON FREQUENCY DISTRIBUTIONS AND COEFFICIENTS DISCUSSED IN CHAPTER THREE

Greater caution will generally be required drawing conclusions based on the Frontline Survey and the In-Person Interviews about topics to be discussed Chapter Three than in using these data sources in subsequent chapters. This is because these instruments gathered data from people still working in these organizations several years after the events we will be discussing, where they were often asked to recall something about those events.   Absent longitudinal research, there is little alternative to using such data if one wishes to study events taking place in the past. However, these data sources create real risks of bias. To illustrate with the frequency distributions reported in Table, there are two reasons to believe these are overestimates of the percentage who actually were earlier supporters of reform. The first is bias from sample attrition. These surveys were limited to people still working in buying offices as of the time of the survey, five years or so after reform had begun. People left these organizations during the intervening years, through retirements and quits, to a significant extent through buyouts offered employees to achieve downsizing, as well (to a very limited extent) through layoffs. One might surmise, though we can't know this for sure since they weren't sampled, that among many reasons for a person to have left was unease about reform. To the extent this is so, samples in these surveys, since they include only current employees, are somewhat biased towards those whose early reaction to reform was positive. This makes the percentage of earlier supporters of reform reported here, based on those still in these organizations, too high.

Two datapoints are available regarding the extent to which critics were differentially likely to leave their organizations after the reform effort began, and they provide somewhat different evidence. In the In-Person Interviews, I asked local office heads to what extent people who took buyouts tended to be opponents of reform. Virtually all responded there was no significant difference, in terms of support for reform, between those staying and those taking buyouts. In the Frontline Survey, I asked respondents to categorize their supervisor's attitude towards reform and also, if they had gotten a new supervisor over the previous two years, to characterize the previous supervisor's attitude. The data show some difference between respondents who still had the same supervisor they had had in l993 and those who had gotten a new one. 46 % of respondents who kept the same supervisor over the whole period reported their supervisor was "enthusiastic" about reform, compared with 42% of those changing supervisors reporting their previous supervisor was enthusiastic; 2% with the same supervisor reported their supervisor was "critical," compared with 8% regarding the previous supervisor.[i] However, these numbers exaggerate differences in attitudes towards reform of people leaving the organization compared to those staying. First, since people generally became more supportive of reform over the period, supervisors who left didn't have a chance to have become more favorable towards reform the way those still there five years later (and about whose <u>current</u> attitude we have information) did; put another way, if we had information about the earlier attitudes of the supervisors who were still there the way we have information about those of supervisors who left (since respondents could only know about the attitude of these previous supervisors at the time), the earlier attitude of those staying would have looked less favorable than that reported, which is based on current attitude.[ii] Also, it is likely any differential propensity of critics to leave was much more pronounced among supervisors than non-supervisors: compared with non-supervisors, supervisors were generally older (and thus more likely to have been eligible for buyouts), were (as we shall see) more likely to have an ideological view of procurement issues and thus find policy changes of which they disapproved more salient, and were more likely to be sensitive to diminished promotion opportunities for reform critics that developed as reform proceeded.

Second, recall questions such as those used here are subject to bias, both random memory lapses and also systematic bias where answers are recalled in the direction of a person's current opinion (current-attitude bias).[iii] So, given that people were on average more favorable towards reform at the time of the survey than when reform began, we can expect they will recall themselves as being more initially favorable to reform than they really had been.[iv] Here, as well, this biases reported percentages upwards, towards showing greater earlier support for reform than actually existed.

Again, we should not exaggerate. While there is surely some current-attitude bias in these responses, evidence from multiple sources in the Frontline Survey suggests that results are not simply artifactual reflections of current attitudes. We can see this, first, by comparing current attitudes to respondent self-reports on TRYBOSS. 36% of respondents rating themselves as positive at the time of the survey to reform (placing themselves over 50 on the procurement reform feeling thermometer) agreed with the TRYBOSS statement, and another 19% characterized themselves as having mixed feelings. So a majority of those reporting themselves as supporting reform at the time of the survey report having initially been skeptics/fencesitters, which, even allowing for random noise, shows current support does not fully determine recall of initial reaction. Second, as we saw earlier in Chapter Three in a different context, people didn't answer two analogous retrospective recall questions in the Frontline Survey – those about initial attitude towards reform and one's initial experience -- even close to identically: they were far more likely to recall having been skeptical when they first tried reform than they were to recall having actually had a bad experience the first time they tried it. If people were simply answering all recall questions in line with current attitudes, one wouldn't expect this divergence. Third, a significant number of respondents circled "don't know" or left these two questions blank (14% and l8% respectively), a higher-than-typical rate for procurement-related questions in the Frontline Survey – for example, 11% answered "don't know" or left blank an analogous question about one's current view of reform, "Acquisition reform has empowered me." This suggests many who couldn't recall their initial reaction to reform avowed this rather than guessing.

Finally in terms of evidence, we analyze both later in Chapter Three, and in Chapter Four, the impact on attitudes towards reform of a scale (called DEFERENCE) measuring the extent to which a respondent was the type of person tending to defer to authority. One might expect this variable to have two effects on support for reform. One would be generally to depress support, since reform sought to encourage people to take initiative themselves, and one would expect those more inclined to defer to authority would be less interested in taking the initiative. The second influence, however, would come from a tendency of more deferent people to accept views of those in authority on the substance of procurement policy. This one might expect to vary as between the beginning of reform and later on: at the beginning, the weight of longstanding authority would have been in support of the traditional system, while by the time of the survey, respondents had experienced a number of years of support for reform by those in authority. And in fact the simple correlation between DEFERENCE and TRYBOSS is .22 (more deferent people were more likely retrospectively to report being less initially supportive of reform, while the correlation between DEFERENCE and PR was -.04 (more deferent people were slightly more likely to be reform supporters as of the time of the survey). If people's answers to TRYBOSS simply reflected their answers to PR, the difference in correlations one would predict on theoretical grounds would not have been observed – one would expect the correlation with TRYBOSS to have been the same as with PR. Similarly, the correlation between a scale measuring venturesomeness/risk-tolerance and TRYBOSS is .28, while the correlation of the same scale with PR is .03. This finding is similarly inconsistent with the view TRYBOSS and PR measured the same thing, but consistent with the view that, when the change was just beginning, less-venturesome people were less willing to try it, while once it became institutionalized, that connection disappeared. These and other theory-consistent observations provide evidence the data do reflect recalled attitudes that are not identical to current ones.[v]

It shouldn't be surprising that answers to these recall questions, while biased, are not meaningless. These questions were about a central feature of their jobs, typically an important feature of a person's life, about which one is likely to remember more accurately than if asked to recall an opinion about foreign policy from five years earlier.[vi] Finally, the finding of the existence of a change vanguard jibes with my own contemporaneous experiences from l993, to be discussed later in Chapter Three, about people expressing dissatisfaction with the traditional system in the time just before reform got launched, although that contemporaneous observation does not reveal the size of the vanguard.

Let us switch to issues with biases in coefficients presented in Table Three. With retrospective recall questions (TRYBOSS and PRVIEW) used to develop the dependent variable in these models, there are two sources of bias, but (fortunately) both affect reported results in a conservative direction, This is so for two reasons. First, with retrospective recall questions, there is greater noise in the data because of random memory errors. The biases p-values upwards, meaning it is possible some coefficients that appear to be statistically insignificant would in fact be significant were there less measurement error.

Current-attitude bias in answering retrospective recall questions – which, as noted earlier, means that frequency distributions exaggerate earlier support for reform – also mean that reported coefficients in these models are too low. The reason this is so is that values for these variables are bounded – for TRYBOSS, for example, there is no available response more pro-reform than to “disagree strongly” (TRYBOSS=5) that one first tried reform only because one’s boss told one to do so. When reform was getting started, there was a real (but now unobserved) distribution of initial attitudes, from strongly critical to strongly supportive, reflected in the survey by TRYBOSS=1 and TRYBOSS=5. When respondents later give a retrospective report of their previous attitude, current-attitude bias leads them on average to recall early attitudes more favorable than attitudes actually were at the time – so, for example, a real “3” will be recalled, say, as a “4.” This means that the observed regression line relating predictor variables to values of the dependent variable would shift upward – for each value of x, respondents would be reporting values of TRYBOSS that were too high. If there were no limit to this shift, the retrospective recall line would remain parallel to the (unobserved) true line, leaving coefficients unaffected, though the intercept of the line would be higher and thus recalled mean values of support for reform too high. However, the variables are bounded -- there is no "pro-reform" answer available more extreme than TRYBOSS=5. This means that those whose true answer from the time would have been TRYBOSS=5 cannot be influenced by current-attitude bias to report an answer higher than TRYBOSS=5 – the boundedness of the variable creates a ceiling for reported values. This means that for respondents whose true value for TRYBOSS was "5," the reported value on the recall question cannot move up in parallel with reported values for respondents whose real value was between l and 4. This prevents the regression line for the recall question from being parallel to the regression line for the true initial reaction, flattening its slope. This makes observed coefficients conservative – they underestimate the strength of the relationship between a predictor and the dependent variable.

The opposite will be occur when the retrospective recall question is a <u>predictor</u> variable rather than a dependent variable – such as, in these models, STRESSWORK5, TIMELY5, GOODK5, or TRYNEW. For the questions where respondents compared a perception of the current situation with that five years earlier (such as the question asking whether the office provided timely service), data are available comparing current with five-year earlier perceptions. In all cases, they show that respondents regarded pre-reform problems with the system as having decreased – they thought, for example, that the system provided more timely service currently than previously and that stress levels had gone down. Here, current-attitude bias would mean that respondents will tend to recall the past as better than it actually was. Thus, a similar inability for people who really had had the most-positive attitude to report a higher recalled attitude makes the slope of the observed regression line steeper than the real line (because of the move towards more positive recalled attitudes, it looks as if a smaller change in x was required to produce a given change in y). However, it is possible that, for questions where the respondent compares past to present, a different recall bias occurs that would counteract this effect.[vii]

In contrast to retrospective recall bias, which tends to make reported results too conservative, problems created by sample attrition (people leaving the organization between l993 and the time the survey was conducted) raise the danger that coefficients might be biased in either direction – they could be too high or too low.

The problem with using respondents in the Frontline Survey to draw conclusions about people in the organization at the beginning of the reform effort in l993 (even removing, as we have, respondents who had not started working for the government in procurement prior to then) is that people who left these organizations after l993 were not a random sample of those in these offices in l993. We could feel confident that departure was truly random only if people were plucked out to leave based on some random procedure such as a lottery. Otherwise, "leavers" will differ from "stayers," even if on no other dimension, in that they chose to leave and thus presumably had some reason(s) to do so. In terms of characteristics we can measure in the Frontline Survey, it is probably true that, as noted earlier, reform critics were somewhat more likely to leave after the change effort got serious than were supporters. Due to "buyouts" offered many older employees, older people were surely more likely to leave than younger ones.[viii] In general, those with better job prospects outside the organization they worked (such as the more highly educated) might be more likely to leave than those with poorer prospects.

If one thinks of the people in the organization as of l993 as the population, and those still there at the time of the survey five or so years later as a sample drawn from this population, then the problem is that non-random attrition means a biased sample of the population has been drawn. A biased sample creates the risk relationships between predictor variables and the dependent variable will be biased – the observed relationship may be stronger than the real relationship, or it may be weaker. The reason is one can't be sure the relationship between a predictor and a dependent variable is the same for those missing from the sample and those in it. For example, imagine a sample of American voters that underrepresented blacks. If among whites the relationship between high income and voting Republican was strong and for blacks it was non-existent, a coefficient for the relationship between income and Republican voting emerging from this biased sample would be too high.[ix] Since coefficients are based on the average relationship between a predictor and a dependent variable, if the sample doesn't reflect the population, and if the relationship between a predictor and dependent variable is different for subgroups in the population, then the average relationship will be biased, just as the mean value for a variable by itself is biased.

The issue of sample attrition, particularly in the context of longitudinal surveys that gather information about a group of people over time – and where some original respondents leave the panel as it continues – has received attention in the econometrics literature. (Heckman l979; a special issue of The Journal of Human Resources in l998 was devoted to this topic.) Econometricians have developed methods, typically quite complex, to estimate corrections for sample attrition, but these require knowledge about characteristics and behavior of “leavers” (so one can compute separate equations to estimate coefficients) and/or the ability to compare the biased sample with another unbiased sample, neither of which is present here.

The extent and direction of any biases in the coefficients depends both on the nature of differences in relationships of predictor variables and the dependent variable between “stayers” and “leavers” (the smaller the differences, the less the bias)[x] and the proportion of “leavers” compared with “stayers” in the original population. The best guess is that the magnitude of coefficient biases here is modest. Overall sample attrition in the Frontline Survey -- the percent of people working for these organizations in l993 had left by the time of the survey -- was xx. (To calculate this figure, I xx) By contrast, after the first five years, sample attrition in the Michigan Panel Study of Income Dynamics, one of the main long-term panel studies subject to economic analysis, was 21.5%, for the Survey of Income and Program Participation 28.6% over its two-year life, for the National Longitudinal Survey of Labor Market Experience 32% for young men and 36% for young women, and for the National Longitudinal Survey on Youth a lower 6.8% for men and 5.2% for women. (Fitzgerald et al: 254; Zabel l998: 480; Falaris and Peters l998: 553, no indication was given of the time period over which Labor Market Experience panel attrition occurred; Macurdy et al: 351) Ridder (l990: 45-46) reports examples of panel studies analyzed in the literature with considerably larger sample attrition rates than these. With this level of sample attrition, overall bias introduced by even relatively significant attrition bias is likely to be quite modest.

Second, although there certainly may be differences between, say, older initial skeptics who stayed in their organizations and older initial skeptics who left, the sample does nonetheless include many people in the categories that disproportionately left the organization, and there is no reason to assume differences between older critics “stayers” and “leavers,” in terms of relationships between predictor variables and the dependent variable, are dramatic.

Overall, studies of sample attrition in the Panel Study of Income Dynamics, the Survey of Income and Program Participation, and the National Longitudinal Survey concluded attrition produced only very modest biases in coefficients (though they did bias mean values for variables), even though attrition was greater among initial sample members with lower socio-economic status. (Fitzgerald et al 1998; Lillard and Panis l998; Zabel l998; Falaris and Peters l998) A study of the National Longitudinal Survey on Youth showed the coefficient relating income to high school graduation produced somewhat higher predicted income for dropouts, compared to the one generated from the Current Population Survey, but the paper concluded it was hard to say the former coefficient was biased. (Macurdy et al l998: 432-34)

Finally, it should be noted that the Frontline Survey, because of the way it was administered (see the Appendix) had an unusually high response rate. Non-response is a problem in many cross-sectional surveys, and introduces the same worries about bias due as does sample attrition over time, so this source of error is likely to be less serious here than in many other cross-sectional surveys.

As discussed earlier in the context of biases in frequency distributions involving some of these variables, absent contemporaneous data collection, we have no choice but to use these data, if we wish to study the phenomena being investigated. This discussion suggests that the overall impact of biases on the results to be reported here is relatively modest.

---

[i] To calculate these figures, respondents who reported having gotten a new supervisor during the previous two years were excluded from the calculations for current supervisor attitude, so one could compare respondent perceptions of the previous supervisor's attitude with perceptions of current supervisor attitude among respondents who had kept the same supervisor over the whole period of time. I also eliminated respondents who had switched offices after l997, so as to remove people who had gotten a new supervisor because they were working at a new place.

[ii] Indeed, many if not most of the supervisors I here call "leavers" may not have left their buying offices at all. In conjunction with the National Performance Review, many previous supervisors were renamed "team leaders," and people's supervisor became a person higher up in the organization; respondents would likely have characterized these situations as ones where they got a "new supervisor." Additionally, some of these organizations routinely make lateral transfers, moving supervisors around different parts of the organization. I am grateful to Edward Elgart, Army Communications and Electronics Command, for enlightening me on these issues.

[iii] Retrospective reports are frequently used in research in areas such as epidemiology, labor market behavior, and consumption habits. They are also used in oral histories and in interview-based qualitative research. There is a fair amount of literature discussing both the strengths and limits of such data. (See generally Pearson et al 1992.) One discussion of the literature concluded that "the findings from the reported studies…(suggest) that memory for everyday autobiographical information can best be characterized as inaccurate in detail but truthful." (Barcklay 1988: 293) A number of studies involving different kinds of data and different disciplinary perspectives establish the presence of current-attitude bias. This was first established in a study (Withey l954) that showed a strong correlation between changes in respondents' incomes between two surveys and the direction of the inaccuracy of recall of income at the earlier period in the second survey: the more, for example, a person's income had increased between the two surveys, the more the person's recall overestimated the earlier income. Goethels and Rickman (l973) exposed high school students to persuasive arguments regarding school busing that caused many of them to change their opinions; after the persuasive efforts, they reported their prior attitudes as being closer to their current ones than a pre-exposure measure had shown them to be. Niemi et al (l980: 64l), examining a panel study of the same respondents in two surveys four years apart, found that (by their clearest measure) only 3.6% of respondents in the later survey recalled that their party identification had been different at the time of the earlier survey, while comparison of the two surveys showed that 22.4% of respondents had in fact changed affiliation – so people's recalled attitudes were biased in the direction of current attitudes. In a similar study (Collins et al 1985: 306) where high school students gave self-reports about alcohol/drug use and, during a re-survey two years later, were also asked to provide retrospective recalls of their use during the early period, the authors performed a regression analysis to see to what extent the recall reports of past use were explained by actual past use a and by current use; they found that actual use explained considerably more variance than current use, but that current use was an important explanatory variable as well. McFarland and Ross (l987) found a similar pattern for recall of attitudes towards a dating partner. To summarize, "(P)eople are cognitive conservatives who bias their memories so as to deny change and maintain temporal consistency and coherence." (Breckler and Greenwald 1986: 134) It may be noted that one political science review article on the evidence on issues of recall bias exaggerates, I believe, the extent

of such bias by referring to any departure in later attitude recall compared to one's answer during the earlier wave of a panel study – including a move, say, from a contemporaneous attitude of "6" on a 7-point scale to a later recall of an attitude of "7" – as biased recall.

It may also be noted that counteracting the general presence of current-attitude bias is the fact that, for opinions that might have been unpopular at the time, current recall may be more accurate than contemporaneous self-reports, if people would have been hesitant to express their true views contemporaneously. (Swan and Newell l998: 125) Arguably, this phenomenon was acting to distort initial self-reports in the McFarland and Ross study of recalled attitudes towards dating partners. This may have been the case for hypothetical contemporaneous self-reports of criticism of the procurement status quo before reform began. In this view, retrospective self-reports might conceivably be more accurate than contemporaneous self-reports, though less accurate than true contemporaneous opinions.

The In-Person Interviews, while also subject to current-attitude bias, are on the whole a more credible source of data than the Frontline Survey, both since people had to express responses in their own words and because recall questions in the interviews were proceeded, as the literature suggests is good practice (see, for example, Loftus and Fathi l985: 282; Fischer and Geiselman l992: 99-102; Fisher and Quigley l992; Schwarz l994), with a specific plea to people to think about the past, not the present, and to think about what was going on in their lives in general, and on the job, in l993 (what Loftus and Fathi call establishing "landmarks"), to focus them on that time period. Thus, these responses may be more valid than answers to retrospective questions in the Frontline Survey, especially since the latter were only a few of many questions in the survey and the respondent is unlikely to have spent any significant time thinking about any individual question.

[iv] As we shall see later in this chapter, a higher percentage of first-line supervisors (where the information was gathered from the Frontline Survey) than of division chiefs (where information was gathered through the In-Person Interviews) reported being earlier supporters of reform. Since current-attitude bias is likely to be less of a problem in the In-Person Interviews than the Frontline Survey (see footnote xx below), this argues for presence of current-attitude bias on these questions in the Frontline Survey. (This is even more the case because, due to differences in questions asked the two surveys, it was somewhat "easier" for a division chief to be characterized as an earlier supporter than for a supervisor.)

[v] Additional evidence will be presented later that is also consistent with the view that current-attitude bias is not an overwhelming problem in these data. Some themes prominent in procurement reform after the effort began were unmentioned by respondents in the In-Person Interviews as criticisms of the traditional system (see footnote xx below). And we will see below (pp.) opposite signs on some variables predicting membership in the change vanguard (versus early recruits) and being an early recruit (versus being a skeptic/fencesitter), which argues these different measures are not artifactual.

[vi] A review article comparing past attitudes to later recall of those attitudes noted (Markus 1986: 29, 36) that recall errors were lower for attitudes that were more salient and that those who most strongly supported women's organizations were most-likely to recall prior attitudes on gender-related issues "Where one is researching areas of high interest one can have that much more confidence that the incidence of misreporting is lower." (Menneer l979: 146) The In-Person Interviews were filled with detailed statements discussing grievances with the pre-1993 system. Responses to a factual question about when the respondent remembered first hearing the Administration was making a push for reform show the ability of most respondents to remember in significant detail specific events from the year reform began. Also, in answering a question in the In-Person Interviews about the first change at their office they associated with procurement reform, virtually all respondents cited changes taking place early in the reform process.

[vii] However, it is also possible that a person's answer to these questions comparing the pre-reform past with the present will be biased in a "good old days"/"bad old days" direction based on his or her current

views of reform, which is the opposite of a retrospective recall bias. In this case, anti-reform people will remember the previous period as “better” (e.g. less stressful) than they actually would have experienced it at the time, so reported results underestimate previous stress, while pro-reform people will remember the previous period as “worse” (e.g. more stressful) than they actually would have experienced it at the time, so reported results overestimate previous stress. – unlike a retrospective recall bias, where reports of the past resemble the present. To the extent that this is occurring, an opposite bias for these coefficients would exist – reported coefficients for these predictor variables would be too low. (A good/bad old days bias could not apply to TRYNEW.)

[viii] Though in the one organization where there were layoffs (RIF’s), younger people rather than older ones would have been so.

[ix] This is why the sample in the Frontline Survey was weighted to reflect more accurately the percentages of Defense and civilian agency procurement employees. See the Appendix.

[x] In some cases, it is possible to develop a prediction about the direction the bias will take. Say that those who were older and more educated/ risk-tolerant (because more-educated people would have better alternative job prospects and less risk-averse people would be more willing to uproot themselves from an existing job) were more likely to leave the organization because of downsizing and other reasons. Among those groups, further assume (which again seems reasonable) that those more initially skeptical of reform were more likely to leave than those more supportive. This means that older people left in the sample would on balance be more initially supportive of reform than older people in the l993 population. Thus, in the sample, we might observe (these numbers are hypothetical) that on average 40 year-olds had a score of 4 on initial support for reform, while 50 year-olds had an average score of 3. However, say the average score of 50 year-olds in the real population from l993, on net, was 2.5, not 3 (because older people tended disproportionately to leave if they were critics). This would mean that the slope observed in the sample was less steep than the one that would have been observed in the population, because in the sample a change in x of l0 years was associated with a change of l.0 in y, while in the population the same change in x would have been associated with a change in y of l.5. Thus, in this example, the coefficient observed in the sample would be too low – it would underestimate the relationship between age and reform skepticism. In other situations, though, the observed coefficient would be too high. For example, if among the more educated, critics tend leave, the mean support for reform among more educated people in the population would have been lower than reform support in the sample – say 3.5 instead of 4. With higher education associated with greater support for reform, this means that the observed coefficient is too high – if the low-educated had a mean reform support score of 3, then a move from low-educated to well-educated in the sample is associated with a change in support of l unit, while in the population the same change in x would have been associated with only a .5 unit change in y, implying the observed coefficient for education is higher than the true coefficient.

To the extent that biased coefficients involve predictor variables correlated with other predictor variables, this introduces biases into coefficients for the other variables as well. So, for example, if the coefficient for education is too high, this suggests observed coefficients for RULES and HIERARCHY are too low, since they are positively correlated with education and the inflated coefficient for education is “stealing” some of their shared variance.