CENSUS ADJUSTMENT: STATISTICAL PROMISE OR STATISTICAL ILLUSION?

David A. Freedman and Kenneth W. Wachter

The census has been taken every ten years since 1790. Counts are used to apportion Congress and redistrict states. Furthermore, census data are the basis for allocating federal tax money to cities and other local governments. For such purposes, the geographical distribution of the population matters rather than counts for the nation as a whole.

Data from 1990 and previous censuses suggested there would be a net undercount in 2000; the undercount would depend on age, race, ethnicity, gender, and—most importantly—geography. This differential undercount, with its implications for sharing power and money, attracted considerable attention in the media and the court-house. There were proposals to adjust the census by statistical methods, but this is advisable only if the adjustment gives a truer picture of the population and its geographical distribution.

The census turns out to be remarkably good, despite the generally bad press reviews. Statistical adjustment is unlikely to improve the accuracy, because adjustment can easily put in more error than it takes out. In this article, we sketch procedures for taking the census, evaluating it, and making adjustments. (A sketch is what you want: detailed descriptions cover thousands of pages.) The article closes with some pointers to the literature, including citations to the main arguments for and against adjustment.

The Census

The census is a sophisticated enterprise whose scale is remarkable. In round numbers, there are 10,000 permanent staff at the Bureau of the Census. Between October 1999 and September 2000, the staff opened 500 field offices, where they hired and trained 500,000 temporary employees. In spring 2000, a media campaign encouraged people to cooperate with the census, and community outreach efforts were targeted at hard-to-count groups.

The population of the United States is about 280 million persons in 120 million housing units, distributed across 7 million "blocks," the smallest pieces of census geography. (In Boston or San Francisco, a block is usually a block; in rural Wyoming, a "block" may cover a lot of pastureland.) Statistics for larger areas like cities, counties, or states are obtained by adding up component blocks.

From the perspective of a census-taker, there are three types of areas to consider. In "city delivery areas" (high-density urban housing with good addresses), the Bureau develops a Master Address File. Questionnaires are mailed to each address in the file. About 70 percent of these questionnaires are filled out and returned by the respondents. Then "Non-Response Followup" procedures go into effect: for instance, census enumerators go out several times and attempt to contact non-responding households, by knocking on doors and working the telephone. City delivery areas include roughly 100 million housing units.

"Update/leave" areas, comprising less than 20 million households, are mainly suburban and have lower population densities; address lists are more difficult to construct. In such areas, the Bureau leaves the census questionnaire with the household while updating the Master Address File. Beyond that, procedures are similar to those in the city delivery areas.

In "update/enumerate" areas, the Bureau tries to enumerate respondents—by interviewing them—as it updates the Master Address File. These areas are mainly rural and post-office addresses are poorly defined, so address lists are quite difficult to construct. (A typical address might be something like Smith, Rural Route #1, south of Willacoochee, GA.) Perhaps a million housing units fall into such areas.

There are also special populations that need to be enumerated—institutional (prisons and the military), as well as non-institutional "group quarters" (for instance, 12 nuns sharing a house in New Orleans are living in group quarters). About 8 million persons fall into these two categories.

Demographic Analysis

Demographic analysis estimates the population using birth certificates, death certificates, and other administrative record systems. The estimates are made for national demographic groups—defined by age, gender, and race (Black and non-Black). Estimates for sub-national geographic areas like states are currently not available. According to demographic analysis, the undercount in 1970 was about 3 percent nationally. In 1980, it was to 2 percent, and the result for 1990 was similar. Demographic analysis reported the undercount for Blacks at about 5 percentage points above non-Blacks, in all three censuses.

Demographic analysis starts from an accounting identity:

$$\begin{aligned}\text{Population} = \text{Births} - \text{Deaths} \\ + \text{Immigration} - \text{Emigration.}\end{aligned}$$

However, data on emigration are incomplete. And there is substantial illegal immigration, which cannot be measured directly. Estimates are made for illegals, but these are (necessarily) somewhat speculative.

Evidence on differential undercounts depends on racial classifications, which may be problematic. Procedures vary widely from one data collection system to another. For the census, race of all household members is reported by the person who fills out the form. In Census 2000, respondents were allowed for the first time to classify themselves into multiple racial categories; this is a good idea from many perspectives, but creates a discontinuity with past data.

On death certificates, race of decedent is often determined by the undertaker. Birth certificates show the race of the mother and (usually) the race of father; procedures for ascertaining race differ from hospital to hospital. A computer algorithm is used to determine race of infant from race of parents. Prior to 1935, many states did not have birth certificate data at all; and the further back in time, the less complete is the system. This makes it harder to estimate the population aged 65 and over. In 2000, demographic analysis estimates the number of such persons starting from Medicare records.

Despite its flaws, demographic analysis has generally been considered to be the best yardstick for measuring census undercounts. Recently, however, proponents of adjustment have favored another procedure, the DSE ("Dual System Estimator").

DSE—Dual System Estimator

The DSE is based on a special sample survey done after the census—a PES ("Post Enumeration Survey"). The PES of 2000 came to be called ACE ("Accuracy and Coverage Evaluation Survey"): acronyms seem to be unstable linguistic compounds. ACE sampled 25,000 blocks, containing 300,000 housing units and 700,000 people. An independent listing is made of the housing units in the sample blocks, and persons in these units are interviewed after the census is complete. This process yields the "P-sample."

The "E-sample" comprises the census records in the same blocks, and the two samples are then matched up against each other. In most cases, a match validates both the census record and the PES record. A P-sample record that does not match to the census may be a "gross omission," that is, a person who should have been counted in the census but was missed. Conversely, a census record that does not match to the P-sample may be an "erroneous enumeration," in other words, a person who got into the census by mistake. For instance, a person can be counted twice in the census—because he sent in two forms. Another person can be counted correctly but assigned to the wrong unit of geography: she is a gross omission in one place and an erroneous enumeration in the other.

Of course, an unmatched P-sample record may just reflect an error in ACE; likewise, an unmatched census record could just mean that the corresponding person was found by the census and missed by ACE. Fieldwork is done to "resolve" the status of some unmatched cases—deciding whether the error should be charged against the census or ACE. Other cases are resolved using computer algorithms. However, even after fieldwork is complete and the computer shuts down, some cases remain unresolved. Such cases are handled by statistical models that fill in the missing data. The number of unresolved cases is relatively small, but it is large enough to have an appreciable influence on the final results.

Movers—people who change address between census day and ACE interview—represent another complication. Unless persons can be correctly identified as movers or non-movers, they cannot be correctly matched. Identification depends on getting accurate information from respondents as to where they were living at the time of the census. The number of movers is relatively small, but they are a large factor in the adjustment equation. More generally, matching records between the ACE and the census becomes problematic if respondents give inaccurate information to the ACE, or the census, or both. Thus, even cases that are resolved though ACE fieldwork and computer operations may be resolved incorrectly. We refer to such errors as "processing error."

The statistical power of the DSE comes from matching, not from counting better. In fact, the E-sample counts came out a bit higher than the P-sample counts, in 1990 and in 2000: the census found more people than the Post Enumeration Survey. As the discussion of processing error shows, however, matching (like so many other things) is easier said than done.

Some persons are missed both by the census and by ACE. Their number is estimated using a statistical model, assuming that ACE is as likely to find people missed by the census as people counted in the census—"the independence assumption." Following this assumption, a gross omission rate estimated from the people found by ACE is extrapolated to the sort of people who are unlikely to be found, although the gross omission rate for the latter group may well be different. Failures in the independence assumption lead to "correlation bias." Data on processing error and correlation bias will be presented later.

Small-Area Estimation

The Bureau divides the population into "post strata" defined by demographic and geographic characteristics. For Census 2000, there were 448 post strata. One post stratum, for example, consisted of Asian male renters age 30–49, living anywhere in the United States. Another post stratum consisted of Blacks age 0–17 (male or female) living in owner-occupied housing in big or medium-size cities with high mail return rates—across the whole country.

Persons in the P-sample are assigned to post strata on the basis of information collected during the ACE interview. (For the E-sample, assignment is based on the census return.) Moreover, each sample person is assigned a "weight." If the Bureau sampled 1 person in 500, each sample person would stand for 500 in the population and be given a weight of 500. The actual sampling plan for ACE is more complex, so different people are assigned different weights, ranging from 10 to 6,000.

To estimate the total number of gross omissions in a post stratum, the Bureau simply adds the weights of all ACE respondents who were identified as (i) gross omissions and (ii) being in the relevant post stratum. To a first approximation, the estimated undercount in a post stratum is the difference between the estimated numbers of gross omissions and erroneous enumerations. The Bureau computes an "adjustment factor"; when multiplied by this factor, the census count for a post stratum equals the estimated true count from the DSE. About two-thirds of the adjustment factors exceed 1: these post strata are estimated to have undercounts. The remaining post strata are estimated to have been overcounted by the census; their adjustment factors are less than 1.

How does the Bureau adjust small areas like blocks, cities, or states? Take any particular area. Each post stratum has some number of persons counted by the census in that area. (The number may be zero.) This census number is multiplied by the adjustment factor for the post stratum. The process is repeated for all post strata, and the adjusted count is obtained by adding the products; complications due to rounding are ignored for now.

The adjustment process makes the "homogeneity assumption," that undercount rates are constant within each post stratum across all geographical units. This is not plausible, and was strongly contradicted by census data on variables related to the undercount. Failures in the homogeneity assumption are termed "heterogeneity."

Ordinarily, samples are used to extrapolate upwards, from the part to the whole. In census adjustment, samples are used to extrapolate sideways, from 25,000 sample blocks to each and every one of the 7 million blocks in the United States. That is where the homogeneity assumption comes into play.

Evaluating Census 2000

We see widespread—although by no means universal—agreement on two chief points. First, Census 2000 succeeded in reducing differential undercounts from their 1990 levels. Second, there remain serious questions about the accuracy of proposed statistical adjustments.

Adjustment faced a new problem with Census 2000. Demographic analysis showed that the census overcounted the population by perhaps 2 million people. Proposed adjustments would have added another 3 million people, making the overcounts even worse. Thus, demographic analysis and ACE pointed in opposite directions. The estimated total national populations are shown in the table.

Demographic analysis	279.6 million
Census 2000	281.4 million
ACE	284.7 million

If demographic analysis is right, there is a census overcount of 0.7 percent. If ACE is right, there is a census undercount of 1.2 percent. Demographic analysis is a particularly valuable benchmark, because it is independent of both the census and the Post Enumeration Survey that underlies proposed adjustments. While demographic analysis is hardly perfect, it is a stretch to blame demographic analysis for the whole of the discrepancy with ACE. Instead, the discrepancy points to undiscovered error in ACE.

Mistakes in statistical adjustments are nothing new. Studies of the 1980 and 1990 data have quantified, at least to some degree, the three main kinds of error: processing error, correlation bias, and heterogeneity. In the face of these errors, it is hard for adjustment to improve on the accuracy of census numbers for states, counties, legislative districts, and smaller areas. Statistical adjustment can easily put in more error than it takes out, because the census is already very accurate.

What went wrong with ACE in 2000? Errors in the statistical operations may from some perspectives have been under better control than they were in 1990. But, it appears, processing error must have been worse in other respects. Research is underway to identify the difficulty. The Bureau is investigating a form of error called "balancing error"—essentially, a mismatch between the levels of effort in detecting gross omissions or erroneous enumerations. We think that troubles also occurred with a new treatment of movers (discussed below) and duplicates. Some 25 million duplicate persons were detected in various stages of the census process, and removed. But how many slipped through?

Besides processing error, correlation bias and heterogeneity are endemic problems that make it extremely difficult for adjustment to improve on the census. Correlation bias is the tendency for people missed in the census to be missed by ACE as well. Correlation bias in 2000 may have amounted, as it did in 1990, to millions of persons. These people cannot be evenly distributed across the country. If their distribution is uneven, the DSE creates a distorted picture of census undercounts.

Heterogeneity means that undercount rates differ from place to place within population groups treated as homogeneous by adjustment. Heterogeneity puts limits on the accuracy of adjustments for areas like states, counties, or legislative districts. Studies of the 1990 data, along with more recent work, show that heterogeneity remains a serious concern.

Missing Data in ACE 2000

Evaluations of the ACE data are ongoing, so conclusions must be tentative. However, there is some information on missing data and on the influence of movers, summarized in the table.

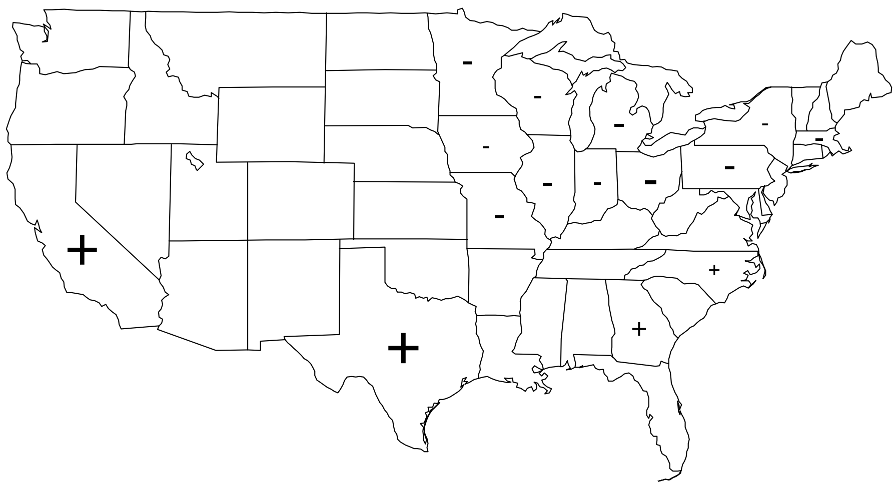
Non-interviews	3–6 million
Imputed match status	3–7 million
Inmovers and outmovers	
Imputed residence status	6 million
Outmovers	9 million
Inmovers	13 million
Mover gross omissions	3 million

These figures are weighted to national totals, and should be compared to (i) a total census population of some 280 million, and (ii) errors in the census that may amount to a few million persons. For some 3 million P-sample persons, a usable interview could not be completed; for 6 million, a household roster as of census day could not be obtained.

Another 3 million persons in the P-sample and 7 million in the E-sample had unresolved match status after fieldwork: were they gross omissions, erroneous enumerations, or what? For 6 million, residence status was indeterminate—where *were* they living on census day? (National totals are obtained by adding up the weights for the corresponding sample people; non-interviews are weighted out of the sample and ignored in the DSE, but we use average weights.)

If the idea is to correct an undercount of a few million in the census, these are serious gaps. Much of the statistical adjustment therefore depends on models used to fill in missing data. Efforts to validate such models remain unconvincing, despite

2000 ACE Adjustment: State Share Changes Exceeding 50 Parts Per Million



some over-enthusiastic claims in the technical and administrative literature. The 2000 adjustment tried to identify both inmovers and outmovers, a departure from past practice. Gross omission rates were computed for the outmovers and applied to the inmovers, although it is not clear why rates are equal—especially within local areas.

For outmovers, information must have been obtained largely from neighbors. Such "proxy responses" are usually thought to be of poor quality, inevitably creating false non-matches and inflating the estimated undercount. As the table shows, movers contribute about 3 million gross omissions (a significant number on the scale of interest) and ACE failed to detect a significant number of outmovers. That is why the number of outmovers is so much less than the number of inmovers. Again, the amount of missing data is small relative to the total population, but large relative to errors that need fixing. The conflict between these two sorts of comparisons is the central difficulty of census adjustment. ACE may have been a great success by the ordinary standards of survey research, but not nearly good enough for adjusting the census.

State Shares

All states would gain population from adjustment. Some, however, gain more than others. In terms of population share, the gains and losses must balance. This subtle point is often overlooked in the political debate. In 2000, even more so than in 1990, share changes were tiny. According to Census 2000, Texas had 7.4094 percent of the population. Adjustment would have given it 7.4524 percent, an increase of $7.4524 - 7.4094 = 0.0430$ percent or 430 parts per million. The next biggest winner was California, at 409 parts per million; third was Georgia, at 88 parts per million.

Ohio would have been the biggest loser, at 241 parts per million; then Michigan, at 162 parts per million. Minnesota came third in this sorry competition, at 152 parts per million. The median change (up or down) is about 28 parts per million. These changes are tiny, and most are easily explained as the result of sampling error in ACE. ("Sampling error" means random error introduced by the luck of the draw in choosing blocks for the ACE sample; you get a few too many blocks of one kind or not quite enough of another: the contrast is with "systematic" or "non-sampling" error like processing error.)

The map shows share changes that exceed 50 parts per million. Share increases are marked "+"; share decreases, "−". The size of the mark corresponds to the size of the change. As the map indicates, adjustment would have moved population share from the Northeast and Midwest to the South and West. This is paradoxical, given the heavy concentrations of minorities in the big cities of the Northeast and Midwest—and political rhetoric contending that the census shortchanges such areas ("statistical grand larceny," according to New York's ex-Mayor Dinkins). One explanation for the paradox is correlation bias. The older urban centers of the Northeast and Midwest may be harder to reach, both for census and for ACE.

The 1990 Adjustment Decision

In July 1991, the Secretary of Commerce declined to adjust Census 1990. At the time, the undercount was estimated by the DSE as 5.3 million persons. Of this, 1.7 million persons were thought to reflect processing errors in the Post Enumeration Survey, rather than census errors. Later research has shown the 1.7 million to be a serious underestimate. Current estimates range from 3.0 million to 4.2 million, with a central value of 3.6 million. (These figures are all nation-wide, and net.) Thus, the bulk of the 1990 adjustment resulted from errors not in the census but in the DSE. Processing errors generally inflate estimated undercounts, and subtracting them leaves a corrected adjustment of 1.7 million. Correlation bias, estimated at 3.0 million, works in the opposite direction, and brings the undercount estimate up to the demographic analysis figure of 4.7 million (see table).

The adjustment	+5.3		
Processing error	−3.6		
Corrected adjustment		+1.7	
Correlation bias		+3.0	
Demographic analysis			+4.7

The message is simple: on the scale of interest, most of the estimated undercount is noise. In 1990, there were many studies on the quality of the adjustment. For 2000, evaluation data are not yet available. However, the Bureau's preliminary estimates, based largely on the experience of 1990, suggest that processing error in ACE contributes about 2 million to the estimated undercount of 3.3 million.

The political debate over adjustment is often framed in terms of sampling: "sampling is scientific." However, from a technical perspective, sampling is not the issue. The crucial questions are about the size of processing errors, and the validity of statistical models for missing data, correlation bias, and homogeneity—in a context where the margin of allowable error is relatively small.

Gross or Net?

Some number of persons were left out of Census 2000 and some were counted in error. Even if ACE had been done with surgical precision, there is no easy way to estimate the size of these two errors separately. Many people were counted a block or two away from where they should have been counted: they are both gross omissions and erroneous enumerations. Many other people were classified as erroneous enumerations because they were counted with insufficient information for matching; they should also come back as gross omissions in the ACE fieldwork. With some rough-and-ready allowances for this sort of double-counting, the Bureau estimated that 6–8 million people were left out of the census while 3–4 million were wrongly included. These are "gross" errors.

Proponents of adjustment are concerned, and legitimately so, about geographical imbalances. Some places may have an excess number of census omissions while other places will have an excess number of erroneous inclusions. Still, adjustment is hardly a panacea. The adjustment mechanism allows cancellation of errors within post strata—the homogeneity assumption at work. In the end, adjustment would have added 4.3 million people nationwide, and subtracted 1.0 million. Much of the gross error is netted out, post stratum by post stratum: the rest is spread uniformly across geography within post strata. Adjustment fixes geographical imbalances in the census only if you buy the ACE fieldwork and the homogeneity assumption.

Proponents of adjustment have also objected to a comparison between undercount estimates (3.3 million in 2000) and estimated processing error (2.0 million), on the grounds that we should not be comparing net errors. We are less sympathetic to this complaint. For most areas with substantial populations—like states—the adjustment is positive, and so are estimates for processing error. Moreover, as noted above, ACE would add 4.3 million in certain post strata and subtract 1.0 million in others. The Bureau's preliminary estimate for processing error has a positive component of 2.6 million and a negative component of 600,000. Any way you slice it, a large part of the adjustment comes about not because of errors in the census, but because of errors in the adjustment process itself.

Loss Function Analysis

Proponents of adjustment often rely on a statistical technique called "loss function analysis." In effect, this technique attempts to make summary estimates of the error levels in the census and the adjustment, generally to the advantage of the latter. However, the apparent gains in accuracy—like the gains from adjustment—tend to be concentrated in a few geographical areas, and heavily influenced by the vagaries of chance. At a deeper level, loss function analysis turns out to depend more on wishful assumptions than on data.

For example, adjustment makes the homogeneity assumption: census errors occur at a uniform rate within post strata across wide stretches of geography. Loss function analysis assumes that and more: not only are census error rates uniform, but so are error rates in ACE. A second example: loss function analysis depends on models for correlation bias, and the Bureau's model assumes there is no correlation bias for females. The idea that only men are hard to reach—for the census and the Post Enumeration Survey—is unlikely on its face. It is also at loggerheads with the data from 1990.

Policy Implications

Decisions not to adjust Census 1980 and Census 1990 were upheld by the legal system. With respect to Census 2000, the courts decided in 1999 that statistical adjustment could not be used for apportionment, that is, the allocation of congressional seats to states. The use of adjustment for redistricting—drawing legislative boundaries within states—was left open.

By April 2001, the Secretary of Commerce had to certify a set of block-level population counts for redistricting. Despite its support for adjustment in 1990, and the advance publicity for ACE, the Bureau advised the Secretary to certify the unadjusted counts, largely because of the conflict between ACE and demographic analysis. Heterogeneity was another concern. The Secretary concurred with the Bureau's recommendation. Adjusted numbers may—or may not—be used to allocate tax moneys: that decision is slated for Fall 2001.

Census 2000 achieved a high level of accuracy. Given that, and given the problems with statistical adjustments, the Secretary's decision to certify the census counts for redistricting was the right decision. So far, this certification has been upheld by the courts, despite challenges by Los Angeles and the Bronx among others.

BIBLIOGRAPHY

Anderson, M. and S. E. Fienberg. *Who Counts? The Politics of Census-Taking in Contemporary America*. New York: Russell Sage Foundation, 1999.

Brown, L. D., M. L. Eaton, D. A. Freedman, S. P. Klein, R. A. Olshen, K. W. Wachter, M. T. Wells, and D. Ylvisaker, "Statistical Controversies in Census 2000," *Jurimetrics*, 39, 1999, pp. 347–75.

Cohen, M. L., A. A. White, and K. F. Rust, eds. *Measuring a Changing Nation: Modern Methods for the 2000 Census*. Washington, D. C.: National Academy Press, 1999.

Darga, K. *Sampling and the Census*. Washington, D.C.: The AEI Press, 1999.

Freedman, D. A., P. B. Stark, and K. W. Wachter, "A Probability Model for Census Adjustment," *Mathematical Population Studies*, 9, 2001, pp. 165–180.

Kass, R. E., ed. "Three Papers on the Census Adjustment," *Statistical Science*, 9, 1994, pp. 457–537.

Prewitt, K. "Accuracy and Coverage Evaluation: Statement on the Feasibility of Using Statistical Methods to Improve the Accuracy of Census 2000," *Federal Register*, 65, 2000, pp. 38373–38398.

Schenker, N., ed. "Special Section on the 1990 Undercount," *Journal of the American Statistical Association*, 88, 1993, pp. 1044–1166.

Singh, M. P., ed. "Special Section on Census Undercount Measurement Methods and Issues," *Survey Methodology*, 18, 1992, pp. 1–154.

Skerry, P. *Counting on the Census? Race, Group Identity, and the Evasion of Politics*. Washington, D. C.: Brookings, 2000.

U. S. Census Bureau. *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy*. With supporting documentation. Washington, D. C., 2001. http://www.census.gov

Wachter, K. W. and D. A. Freedman. "The Fifth Cell," *Evaluation Review*, 24, 2000, pp. 191–211.

Wachter, K. W. and D. A. Freedman. "Measuring Local Heterogeneity with 1990 U. S. Census Data," *Demographic Research*, an on-line journal of the Max Planck Institute. Volume 3, Article 10, 2000. http://www.demographic-research.org/Volumes/Vol3/10/

David A. Freedman is Professor of Statistics and Mathematics, University of California, Berkeley, CA 94720-3860. His research interests are in the foundations of statistics, modeling, and policy analysis.

Kenneth W. Wachter is Professor of Demography and Statistics, University of California, Berkeley CA 94720-3860. His research interests include mathematical demography, demographic heterogeneity, computer microsimulation, kinship forecasting, and statistical astronomy.

Both authors testified for the United States against adjustment, in the Census cases of 1980 and 1990. They have also testified in Congressional hearings, and consulted for the Department of Commerce on the adjustment decision.