Results of Quality Assurance on the Accuracy and Coverage Evaluation Matching Operations

Tamara Adams, Rosemary Byrne, and Magdalena Ramos[1]
Decennial Statistical Studies Division
Bureau of the Census
Washington, D.C. 20233

Keywords: Accuracy and Coverage Evaluation, Census 2000, quality assurance

I. Introduction

As part of the Census 2000, the Census Bureau conducted the Accuracy and Coverage Evaluation (A.C.E.). First, the census completed an enumeration of all people. Then, the A.C.E. interviewed a sample of 314,649 housing units in 11,802 sample clusters in the 50 states, the District of Columbia, and Puerto Rico. We then matched the A.C.E. people in A.C.E. clusters (P-sample) to the census people in A.C.E. clusters (E-sample). The results of the matching was used in the dual system estimation. The P-sample was used to estimate the number of people missed in the census; the E-sample was used to estimate the number of people erroneously enumerated in the census.

We performed the matching in several phases:

- **Computer Matching** – We computer matched the P-sample to the census people within each cluster.
- **Before Followup Matching** – Clerical matchers review the results of computer matching to find additional matches and to code duplicates records within the P-sample and within the census.
- **Person Followup** – Interviewers followup cases which needed additional information
- **After Followup** – Clerical matchers reviewed batches of person followup forms and code each person followed up as correctly counted within the cluster or incorrectly counted within the cluster.

Clerical matchers use an automated system to review the person records, including name and demographics, to attempt to match and code the records during the before followup stage and the after followup stage.

To control the matching operations, we developed a quality assurance (QA) plan to minimize matching error while allowing a timely flow of work.

II. Quality Assurance Plan for Person Matching

While developing the QA plan, we decided to use a three-tiered dependent review of person records:

- **Clerks** – The clerks reviewed a workunit first and coded all records not matched in computer matching.
- **Technicians** – The technicians reviewed the clerks' work, checking for errors. However, the technicians did not have to recode each record reviewed if they agreed with the code.
- **Analysts** – The analysts reviewed the technicians' work and did not have to recode each record reviewed if they agreed with the code.

This allowed a faster review than an independent rework and adjudication because the successive levels did not have to recode each record, as in 1990. In addition, it minimized error because each successive level had more technical expertise and training than the previous level.

Within each phase, matchers reviewed person records at the household level. Each phase has several stages for QA purposes. Before followup has three stages; after followup has five stages. However, the workunits varied with each stage:

- **Before Followup** – Matchers reviewed records within the A.C.E. sample clusters
- **After Followup** – In the first three stages of this operation, matchers reviewed batches of followup forms. In the last two stages, matchers reviewed records within A.C.E. sample clusters as in before followup.

The QA plan relied on change rates to assess the quality of matchers' work. The matching system calculated the change rates using prespecified algorithms that compared the matchers' match codes and determined significant changes made by the reviewer. Individual code changes did not always indicate errors. Additional experience and training may have led a matcher at a higher level to code

[1]This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

a record differently. A significant change for each stage is defined below:

- Before followup – A change from a code that would have prevented a person record from being assigned to person followup.
- After followup – A change from a code that would have caused a person record to be assigned the wrong enumeration, match, or residence status.

We designed the QA methodology to target both matchers who required more consistent review and records that needed a higher level review. We used several steps to ensure high quality review:

- **Prequalification** – Clerks and technicians reviewed a preselected set of training clusters before production. A computer program compared their answers to an answer key. Managers then selected the matchers with change rates below four percent to begin review production under a sample review. Technicians reviewed batches in a similar fashion for after followup prequalification.
- **Initial Sampling Decision** – For matchers who were not prequalified, the matching system calculated change rates after a user worked a prespecified number of records. If the change rate was less than four percent, then the system placed the user into sampling.
- **Reassessment** – After the initial decision, the system reassessed the user after every 50 records worked that were reviewed in clusters selected for a 100 percent review. Based on a change rate cutoff of four percent, the system placed matchers with a higher change rate into 100 percent review and placed the matchers with a change rate of less than four percent into sampling mode.
- **Targeting by System** – The system also checked for certain predefined situations and, if present, automatically routed the cluster or batch to a higher level matcher for review.
- **Targeting by User** – The system allowed matchers to flag difficult records for the next level of review.
- **Data Edits** – The system assessed the matchers' coding during review and at closeout of a cluster or batch. The user could not assign invalid match codes or leave a record uncoded.

The QA plan allowed us to control the data quality that resulted from the person matching operations. The clusters and batches selected in a sample review allowed us to control for random errors, while those that were targeted allowed us to control for difficult situations. In addition, we constantly reassessed matchers throughout the matching process. The reassessment allowed for an improvement over time as a user climbs the learning curve and also accounted for matchers who did not perform up to standards throughout the entire matching operation.

III. Assumptions and Limitations

The QA plan had several assumptions:

- **Change Rate** – The change rate always overestimated the true error rate. We cannot assess the true amount of work a higher level user reviewed due to the targeted clusters. In these clusters, the higher level matchers only had to recode certain types of records. However, the matchers reviewed records above and beyond the required amount. Therefore, we compensate by calculating our overall change rates (below) using three models to more accurately reflect the production efforts.

- **No QA on the Analyst Level** – We consider the analysts to be expert matchers. Due to their extensive training and experience, we assume that analysts have no error.

- **Out-of-Scope Stages** – The last two stages of after followup were cluster-based, rather than batch-based. These two stages are disregarded for QA purposes. The first cluster review stage, performed by technicians, consisted of a targeted review of a certain type of record with very defined coding rules. The second review stage, performed by analysts, consisted of a full cluster review of targeted clusters. We do not consider code changes in these clusters as errors in the batch work due to the availabililty of additional cluster-based information in the cluster review stages. Since we have additional cluster-based information in the cluster review stages not available in the batch stages, we do not consider code changes in these clusters as errors.

- **Comparison Between Phases** – Due to the different nature of the before followup work and the after followup work, change rates of either phase cannot be used to assess the QA program for the other phase.

In addition, we had limitations on the data presented in this paper:

- **Cluster Review Stages** – As stated above, these two stages are not included in any data presented in this paper.

- **Computer Matching** – We computer matched 69.6 percent of the P-sample and 64.4 percent of the E-sample in the Computer Matching Phase using the Census Statistical Research Division Record Linkage System. The computer matcher assigned cutoffs very conservatively. Numerous studies over the years have shown that this operation was virtually error free (e.g., there were insignificant numbers of false matches). So, outgoing change rates given below only apply to about 30% of all records (those that were not computer matched).

IV. Methodology of Outgoing Quality Rate Calculations

We calculated outgoing quality rates for each stage–first, individual user rates for each user. Then, we modeled the overall outgoing quality rate for each stage. The calcuations for before followup matching and after followup coding are separate calculations, but are performed using the same algorithms.

Individual User Change Rates

- **Individual user change rates for technicians –** We calculate individual change rates for each technician by dividing the records changed by the records checked at a higher level.

- **Individual user change rates for clerks** – For the clerks, we calculate individual change rates by dividing the records changed at a higher level by the records checked at a higher level, with an adjustment factor. We also calculate an adjusted overall technician change rate to compensate for the changes the technicians may have missed. We use this factor to adjust the clerks' individual change rates. We calculate each of these rates for sampled clusters/batches only.

Estimating the Overall Outgoing Quality Rate

We used three different models to estimate the overall change rates for clerks and technicians. For any given user, we classified records four ways:

- Randomly sampled for review (X)
- Not sampled for higher review but part of a cluster that a higher level user reviewed and coded (Y')
- Not sampled for higher review but part of a cluster that a higher level user reviewed and did not code (Y")
- Not sampled for higher review and not worked at a higher level (Z)

From the records in X, we had individual change rates, generalized here as p_x, for a given user as discussed above. The sum of Y' and Y" was Y, a cluster or batch that the system did not sample for higher review but that a higher level matcher worked. Using the proportion $p_x=errU_i$, we estimated the overall change rate for each type of user with the following equation:

Before Followup Overall Change Rate=

$$\frac{\left(\sum_{i}^{n} errU_i * recsU_i\right)}{\left(\sum_{i}^{n} recsU_i\right)}$$

$errU_i$ = Before Followup individual change rate for user U_i
$recsU_i$ = records worked by user U_i
n = number of users in a given stage

The overall change rates are calculated separately for each stage and each type of user (clerk or technician). For each model, the $recsU_i$ varies depending on the assumptions in the model.

We estimate the outgoing quality as follows:

Outgoing quality of the phase
= (1 - *overall change rate*)*100

The outgoing quality is also calculated separately for each stage and each type of user.

For the clerk level, records were considered part of Y if a technician reviewed the cluster or batch, but the workunit was not sampled for QA. For the technician level, records were considered part of Y if an analyst reviewed the cluster or batch, but the workunit was not sampled for QA.

Model 1:

- Assumptions: Y and Y' were random. That is, the clusters/batches that were not sampled for a higher review were chosen randomly and the records that were recoded were chosen randomly.
- Inference: The individual change rate estimated the changes present in Y" and Z.

Model 2:

- Assumptions: Y was random and Y' was not random. That is, the clusters/batches that were not sampled for a higher review were chosen randomly, but the records that were recoded were targeted because they were incorrect.

- Inference: The individual change rate estimated the changes present in Z. There were no remaining defects remaining in Y"because all of the changes (Y') were corrected.

Model 3:

- Assumptions: Y was not random. That is, the clusters/batches that were not sampled for a higher review were not chosen randomly.

- Inference: The individual change rate estimated the changes present in Y and Z. We know that we corrected Y' changes and removed those cases from our change count.

We calculate the AFU Technician change rates, both individual and overall, and the outgoing quality in the AFU Technician stage, the same way as the BFU Technician change rates only using records from the AFU Technician and AFU Analyst stages.

We calculate the AFU Clerk change rates, both individual and overall, and the outgoing quality in the AFU Technician stage, the same way as the BFU Clerk change rates, again considering records reviewed by technicians differently than those reviewed by analysts.

V. Results

Table 2 shows the coding changes in sampled clusters/batches by level of highest review. Records worked in the clerical matching operation only appear in one row. That is, if a clerk and technician both worked the record in BFU in sampled clusters, this record appears only in the BFU Technician row.

Table 3 shows the overall technician and clerk change rates by stage. Only records to be worked by a clerical user are included; all records that were computer matched and not reviewed are excluded.

Figures 1-4 show the individual user change rates for each stage and type of user. For each figure, the solid line shows the individual change rate for the given stage/type of user in the sampled clusters/batches and the dashed line shows the corresponding individual contribution to the overall change rate calculated using model 1.

Table 1–Overall Outgoing Quality Rate Models

Model	Y (clusters/batches reviewed, not selected for QA)	Y' (records in clusters/batches not selected for QA, but coded)	Estimation Formula
1	Random	Random	$\frac{p_x*(Y''+Z)}{X+Y+Z}$
2	Random	Not Random	$\frac{p_x*(Z)}{X+Y+Z}$
3	Not Random	n/a	$\frac{[p_x*(Y+Z)]-Y'}{X+Y+Z}$

Table 2 – Records Reviewed in QA Clusters

Stage	Technician Changes	Total records reviewed by Technicians	Analyst Changes	Total records reviewed by Analysts
BFU Clerk	4,859	150,353	1,246	46,315
BFU Technician	n/a	n/a	753	
AFU Clerk	2,060	86,204	657	34,997
AFU Technician	n/a	n/a	847	
Total (n)	6,919	236,557	3,503	81,312

VI. Conclusions

- Matching QA was successful at minimizing errors – The outgoing quality rates are shown above in Table 2. All outgoing quality rates are higher than 99%.

- Matching QA minimized errors at a user level – As we can see from figures 1-4, the QA system targeted matchers with higher change rates; the system kept more of these matchers' clusters and batches in full review (these matchers are represented on the far left side of each graph). The matchers with the highest change rates had the greatest difference between their change rates (dark line) and residual change rates (dashed line). Therefore, the QA system equalized the residual change rate of all matchers; all matchers contributed similar residual change, regardless of the change rate in their sampled clusters/batches.

VII. References

Byrne, R. (May 2000), "MaRCS Specifications for Quality Assurance (QA) for Person Matching", DSSD Census 2000 Procedures and Operations Memorandum Series Chapter S-QA-13.

Childers, D. (January 2001), "Accuracy and Coverage Evaluation: The Design Document", DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1.

Childers, D. , Feldpausch, R., Adams, T. (February 2001), "Accuracy and Coverage Evaluation: Person Matching and Follow-up Results, DSSD Census 2000 Procedures and Operations Memorandum Series, B-6*.

Table 3 – Overall Change Rate and Outgoing Quality Rate by User Type and Stage

Stage	Model 1		Model 2		Model 3	
	Overall Change Rate	Outgoing Quality Rate	Overall Change Rate	Outgoing Quality Rate	Overall Change Rate	Outgoing Quality Rate
BFU Clerk	0.59%	99.41	0.52%	99.48	0.44%	99.56
BFU Technician	0.23%	99.77	0.22%	99.78	0.20%	99.80
BFU Analyst	0.00%	100	0.00%	100	0.00%	100
AFU Clerk	0.95%	99.05	0.11%	99.89	0.30%	99.70
AFU Technician	0.71%	99.29	0.13%	99.87	0.24%	99.76
AFU Analyst	0.00%	100	0.00%	100	0.00%	100

Figure 1--Before Followup Clerk Change Rates

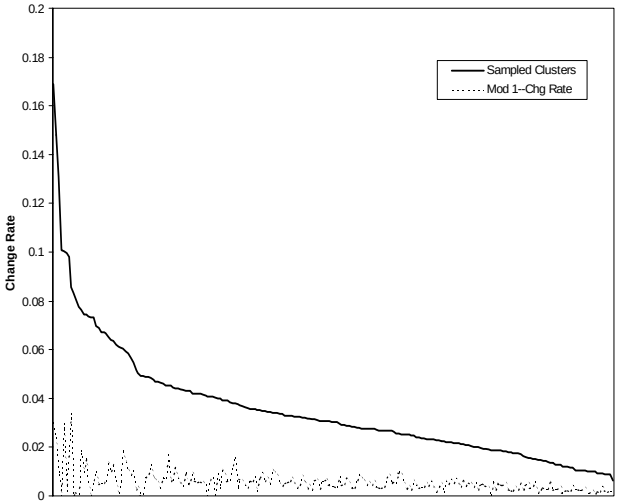

Figure 2--Before Followup Technician Change Rates

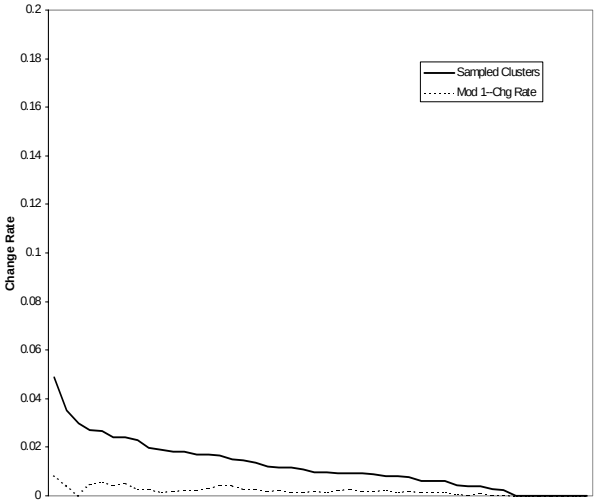

Figure 3--After Followup Clerk Change Rates

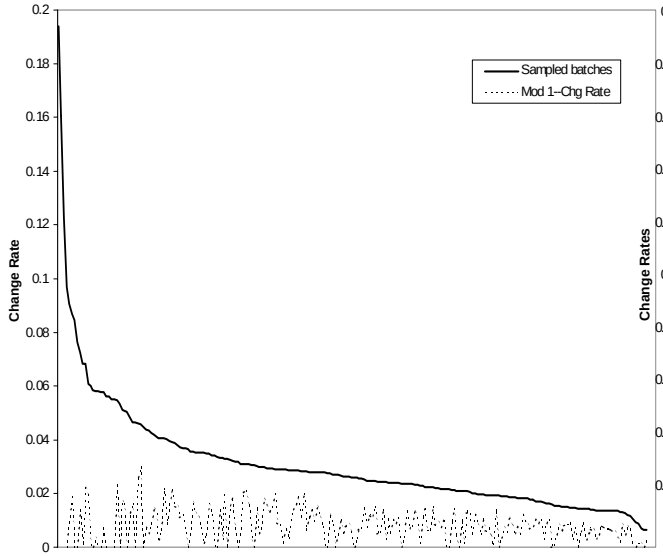

Figure 4--After Followup Technician Change Rates