BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-91/05

THE STATISTICAL RESEARCH DIVISION'S SUPPORT OF QUALITY ASSURANCE IN DECENNIAL CENSUS GEOGRAPHIC OPERATIONS

by

Brian Richards
Statistical Research Division
U.S. Bureau of the Census
Washington, DC 20233

Report issued: June 11, 1991

THE STATISTICAL RESEARCH DIVISION'S SUPPORT OF QUALITY ASSURANCE IN DECENNIAL CENSUS GEOGRAPHIC OPERATIONS

Brian Richards
Statistical Research Division
June 11, 1991

I. OVERVIEW

Geography - related operations were an integral part of the preparations for the 1990 Decennial Census. The Statistical Research Division (SRD) was responsible for the design of the quality assurance program for many of the geographic operations. Much of our responsibilities involved devising quality control (QC) sampling plans to be used in assuring a desired level of accuracy for operations of interest. SRD also observed these geographic operations at field sites for the purpose of determining whether the QC program was being carried out in accordance with agreed-upon specifications.

The two major geographic operations in which SRD was involved were:

(1) **The 1990 Tape Address Register (TAR) Clerical Geocoding Operations.** The operations involved the assignment of correct geographic codes to addresses. Geocoding allowed addresses to be matched to their physical location.
(2) **Geographic Update System (GUS) Operations.** GUS was a series of operations which involved geographic updates to the TIGER file requiring additions, deletions, and modifications to features, addresses and geographic codes.

Section II describes our involvement in the Clerical Geocoding operation. Section III describes our involvement in the GUS operations. The attachment to this report contains a listing of all relevant memorandums relating to SRD's involvement in these operations. There were several other geography — related operations in which SRD was also involved, although to a much lesser extent. These operations are as follows:

(1) **The keying of the geocodes assigned during the TAR Clerical Geocoding Operation into the appropriate data base.**
(2) **1990 Collection Geography Insertion data entered into the TIGER file.** Codes representing political and statistical areas (such as census tracts or block groups) were added, deleted or revised in the TIGER file.
(3) **The Address Insertion Project (AIP).** The AIP involved clerks inserting ZIP Codes and Address Ranges (corresponding to annotated segments on map sheets) into the TIGER file.
(4) **The annotation of Tract Codes onto appropriate maps for use in the Post Enumeration Survey (PES).** The operation involved clerks annotating each tract with the proper socio - economic stratum codes.

There is a supplement to this technical report which contains all of the background memorandums associated with SRD's support of these operations. This supplement is entitled "Supporting Memorandums - Quality Assurance associated with Decennial Census Geographic Operations", and consists of specifications, trip reports, observations, and evaluations. This supplement should be consulted if additional details on findings discussed in this report are desired.

II. THE 1990 TAPE ADDRESS REGISTER (TAR) CLERICAL GEOCODING OPERATION

In 1988, the Geography Division (GEO) matched vendor address files covering the Tape Address Register (TAR) areas against the TAR geography in the TIGER file, and geographically coded (geocoded) by machine all matched addresses. The addresses that failed to match were uncoded and referred to as unmatched addresses. Slightly more than one-half of the addresses in the nation (58,000,000 out of 104,000,000) fell within TAR areas. These areas contained city-type mail delivery. About 14,000,000 addresses were unmatched during the initial machine-match and thus were subject to clerical resolution. Clerical resolution of these unmatched addresses was performed at the 12 Regional Census Centers (RCCs), as well as several other sites around the country, from May 1988 through January 1989. The clerical geocoding operation involved assigning geographic classification codes (district office [DO], address register area [ARA], block) to the unmatched addresses listed for geocoding resolution.

The 58,000,000 addresses in TAR were contained in 345 large cities, thus these cities were designated as TAR areas. To facilitate clerical operations, Coding Areas were created. There was usually a one-to-one correspondence between TAR areas and Coding Areas, although in a few places adjacent TAR areas had common address source materials. In these areas a Coding Area was represented by more than one TAR area, thus there were only 340 Coding Areas. Within a Coding Area, the addresses were grouped by cluster. A cluster consisted of a blockside of addresses with the same feature identifier (street name), the same ZIP Code, the same address parity (odd or even), and the same hundred range. Clerks working in the field offices used geographic sources such as local maps and ZIP Code directories in their attempt to geocode the unmatched clusters.

For the purpose of processing and QC, the clusters were grouped into work units containing 1 to 125 clusters. SRD designed an independent QC system, with collaboration from GEO and Field Division (FLD). The plan called for a sample of clusters to be randomly chosen from each work unit. For each sampled cluster, a precode clerk determined the appropriate actions (before processing occurred) to be taken in assigning the geocodes, and documented these actions. An adjudicator compared the actions taken by the precode and processing clerks and tallied the number of critical processing errors found, which was then used as a basis for the work unit decision (pass or fail). Specific details of the independent QC program can be found in the January 12, 1988 memorandum which is included in the supplement to this report.

The independent QC system was discontinued in October 1988. The decision to discontinue the independent QC system was made because it appeared very unlikely that the completion date for TAR Clerical Geocoding would be met if the existing QC plan was to continue. This decision was agreed upon by Decennial Planning (DPLD), FLD and GEO Divisions. They stated that they were willing to accept the potential risk of lower quality in exchange for a higher coding rate that met the deadline for deliverables.

While the independent QC system was in place, QC worksheets were being completed for each processed work unit. The QC worksheet contained information such as number of clusters in the work unit, the number of clusters sampled, and the number of clusters determined to be erroneously geocoded (along with an error category for each error). Appropriate data from each of these QC worksheets were entered into a computer data base back at headquarters. This enabled the data to be analyzed in various ways, allowing for comparisons of different categories of variables such as RCC, Coding Area, and error type.

The overall error rate, based on QC data from 2700 work units, was 8.6%. The independent QC system included a mechanism for rework and then rechecking by the adjudicator for failed work units of clusters. In theory, this rework mechanism should have improved the "outgoing quality" rate to the agreed upon target of 95% accuracy (i.e., an error rate of 5% or less). One has to realize that the attainment of a 95% accuracy level was based on several unrealistic assumptions. For example, it was assumed that the adjudicator detected (and thus corrected) all errors contained in the sampled clusters. In addition, it was assumed that failed work units were totally correct after rework and rechecking by the adjudicator. Because of the probable inability to meet these assumptions, it is likely that the outgoing quality was less than the desired target of 95% accuracy.

The December 13, 1988 and October 4, 1990 memorandums contained in the Supplement included a number of graphs and tables resulting from the analysis of the QC data. One of them, Figure 1, is included in this paper, and displays a graphical representation of the estimated error rates across the RCCs.

During SRD's review of the QC worksheets, some problems and inefficiencies were observed with regard to the execution of the QC program. Continuing recurrence of critical errors was identified in many Coding Areas, which caused excessive time to be spent on rework and rechecking by adjudicators. These recurring errors appeared to be caused by a lack of effective feedback and/or retraining given to processing clerks when critical errors were committed.

The estimated error rate of 8.6% stated previously should be further clarified. This estimate was calculated from data generated while the formal independent QC program was in place. It is likely that the geocodes assigned after the formal QC program was discontinued will be less accurate than geocodes assigned while the formal QC program was in place. History has shown us that when QC is removed from an operation, the quality of the data usually decreased significantly. In addition, one should also realize that many of the "difficult" areas to geocode (such as New York, Philadelphia, and Los Angeles) were processed at the end of the operation, when there was no formal QC in place.

As discussed above, the formal independent QC program was discontinued in October 1988. One negative consequence of this was the resulting inability to measure the accuracy rate of the operation. Therefore, a special purpose evaluation to assess the accuracy of the 1990 TAR Clerical Geocoding operation was planned in early 1989. It was believed that an estimate of the error rate, along with a characterization of the types of errors, would help in the planning of future operations.

A sample of clusters from the operation was chosen to represent the evaluation. The evaluation called for the sampled clusters to be independently geocoded, and the results compared to the geocodes assigned during clerical processing. It was agreed that the geographers in the RCCs were the best qualified (and most logical) candidates to perform the assignment of geocodes for the sampled clusters, given the timing constraints and logistics involved. Since it was believed that their work would be highly accurate, it was planned to use their geocode assignments as the basis for the evaluation.

The geographers in the RCCs performed the assignment of geocodes for the sampled clusters in March 1989. Because of circumstances specific to the RCCs, someone other than the geographer performed the assignment of geocodes for the evaluation in some instances. SRD performed their analysis of the resulting data in March and April, 1989. A total of 1255 clusters were involved in the analysis. It was not possible to formally publish estimates relating to the accuracy of the operation, since there was evidence that invalidated the assumptions crucial to the evaluation.

A REX geocoding evaluation is being planned for the years 1991-93. Representatives of various divisions have been meeting monthly since April 1990. SRD is among those divisions which are playing an active role in planning the evaluation. All addresses nationwide (with a few very minor exceptions) will be included in the frame from which the sample will be selected, not just those addresses subject to TAR clerical resolution. The objectives of the evaluation are:

(1) Estimate the accuracy of geocoding nationwide.
(2) Determine the source of geocodes assigned.
(3) Determine the cause of error in incorrect geocodes.
(4) Determine the possible effects of the incorrectly-assigned geocode.

III. GEOGRAPHIC UPDATE SYSTEM (GUS) OPERATIONS

The Geographic Update System (GUS) is an interactive update system that is used to perform a series of operations involving geographic updates to the TIGER files requiring additions, deletions, and modifications to features, addresses, and geographic codes. The GUS operations began in the latter part of 1988. Most operations were completed in 1990, but a few have continued into 1991. The operations are being conducted at the 12 RCCs; at the Data Preparation Division (DPD) in Jeffersonville, IN; and at the Digital Files Branch (GEO).

The GUS operations being carried out include (some have already been concluded):

(1) The updating of TIGER based on:

(i) Geographic updates derived from annotations made to ARA maps during the 1990 TAR Clerical Geocoding operation.
(ii) Updates from the 1988 and 1990 Boundary and Annexation Survey (BAS).
(iii) Geographic updates placed on maps by enumerators during Prelist and Precanvass.
(iv) Geographic updates placed on maps during the Cycle I Block Split Operation.
(v) Updates received from local officials.

(2) The insertion of appropriate codes into the TIGER file associated with:

(i) Voting Districts (VTDs).
(ii) Census Designated Places (CDPs).
(iii) Congressional Districts (CDs).

(3) The digitizing of Urbanized Area boundaries into the TIGER file.

Clerks working out of the field offices typically conduct these updates at digitizing stations. Digitizing is the process which converts individual line segments on maps into computer readable form. These clerks, who are often referred to as digitizing operators, then input the geographic updates (that they see annotated on the paper maps) into the appropriate TIGER data base.

In accordance with the QC Sampling Plan, map sheets were grouped into work units containing ten or fewer map sheets. One or more processed map sheets were chosen from each work unit for QC, depending on the number of geographic updates contained on the map sheets(s). The number of critical errors was tallied by the QC operator. Acceptance criteria (relating to the number of errors identified) were applied to each work unit, and rework was performed when necessary. Additional details of the QC sampling plan may be found in the April 20, 1989 memorandum which is included in the supplement to this report.

QC forms were completed at each site documenting the results of the QC program. The level of conformance to the QC plan varied greatly from site to site. Problems such as not reworking failed work units were identified at some sites. It was impossible to estimate the accuracy of these operations since (1) appropriate QC data were sometimes not documented, and (2) there was no program in place allowing for input of QC data into a data base from which analyses could be performed and inferences made.

A training session explaining the quality assurance program used in GUS operations was presented at each of the 14 field sites. The training session was developed by Brian Richards of SRD, with collaboration from Laura Lunsford and Joseph Prchlik of GEO. The training session was entitled "Overview of Quality Assurance Programs" and included topics such as:

(1) Formation of work units.
(2) The QC sampling plan.
(3) Using the acceptance criteria tables.
(4) Methodologies for tallying actions and errors.
(5) Rework and rechecking requirements.
(6) How to complete QC forms.
(7) Discussion of acceptance sampling and process control.

In addition to explaining basic acceptance sampling concepts relating to the QC sampling plan, the training session also emphasized the concepts of quality planning and process control. The training session was presented at each of the 14 field sites between July 1989 and June 1990.

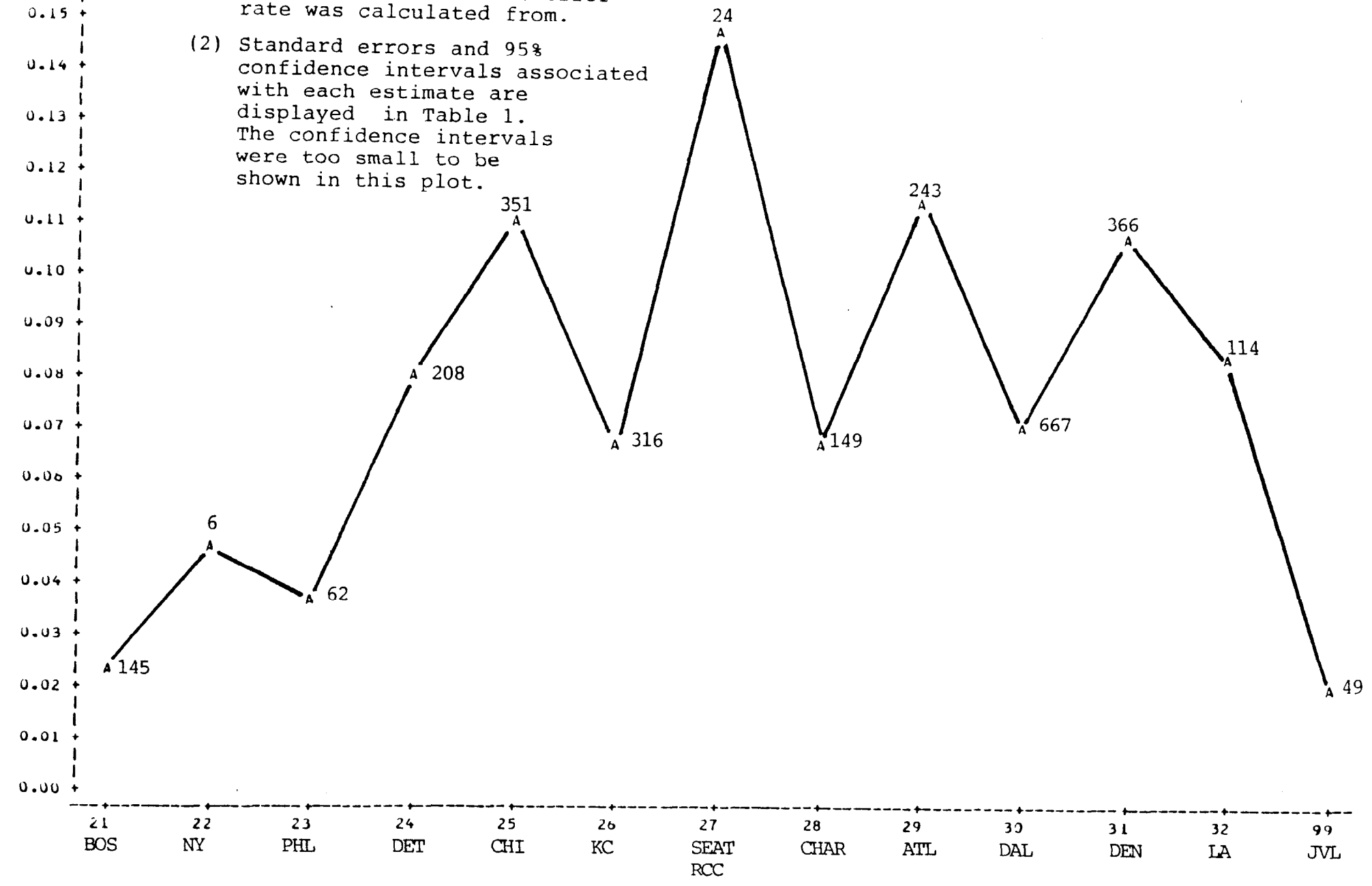
A number of trip reports are contained in the supplement to this report describing details of the training session presentations, as well as observations resulting from field site visits.

Recent discussions regarding future plans for GUS quality assurance revealed a desire within SRD and GEO to automate the quality assurance program applied to geographic update operations associated with the 2000 Decennial Census. The goal is to have a QC sampling plan (to check the accuracy of geographic operations) which is both automated and independent. This desire to improve the quality assurance program at this early date in the next 10 year census cycle is seen as a very positive development, and is consistent with the philosophy of quality planning (which should be a primary objective in planning for the 2000 Decennial Census). Discussions and planning are currently under way to work towards this goal over the next ten years.

FIGURE 1 Plot of Error Rate vs. RCC

NOTES: (1) The number displayed by each point represents the number of work units that the error rate was calculated from.

(2) Standard errors and 95% confidence intervals associated with each estimate are displayed in Table 1. The confidence intervals were too small to be shown in this plot.



ATTACHMENT

LISTING OF
MEMORANDA RELATING TO THE STATISTICAL RESEARCH DIVISION'S SUPPORT OF QUALITY ASSURANCE IN DECENNIAL CENSUS GEOGRAPHIC OPERATIONS

I. Memoranda associated with 1990 TAR Clerical Geocoding operations:

(1) "Quality Control Plan for Field Geocoding for the 1990 Decennial Census"
Kirk M. Wolter to Robert W. Marx
January 12, 1988

(2) "Observation of Geocoding Activities for 1990 Decennial Census in Atlanta Regional Census Center"
Kirk M. Wolter to Stanley D. Matchett
June 16, 1988

(3) "Observation of Geographical Operations for 1990 Decennial Census in Denver Regional Census Center"
Nash J. Monsour to Stanley D. Matchett
September 13, 1988

(4) "Observations of Quality Control Program associated with 1990 TAR Geocoding Operations"
Lawrence R. Ernst to Stanley D. Matchett
December 13, 1988

(5) "Evaluation Study for 1990 TAR Geocoding Operations" (Specifications)
Nash J. Monsour to Susan M. Miskura
March 9, 1989

(6) "Evaluation Study for 1990 TAR Geocoding Operations" (Results)
Lawrence R. Ernst to Susan M. Miskura
June 7, 1989

(7) "Final Observations - Quality Control Program associated with 1990 TAR Clerical Geocoding Operations"
Robert D. Tortora to Stanley D. Matchett
October 4, 1990

II. Memoranda associated with Geographic Update System (GUS) operations:

(1) "Observation of Geographic Update System operations at Chicago and Kansas City Regional Census Centers"
Lawrence R. Ernst to Stanley D. Matchett
January 31, 1989

(2) "Observation of Geographic Update System operations at the Philadelphia Regional Census Center"
Lawrence R. Ernst to Stanley D. Matchett
February 3, 1989

(3) "Quality Control Sampling Plan for Geographic Update System Operations"
Nash J. Monsour to Robert W. Marx
April 20, 1989

(4) "Observation of Boundary and Annexation Survey (BAS) Operations in Data Preparation Division"
Lawrence R. Ernst to Joseph S. Harris
June 7, 1989

(5) "Observation of Boundary and Annexation Survey (BAS) Digitizing Operations in Chicago and Charlotte Regional Census Centers"
Lawrence R. Ernst to Stanley D. Matchett
July 10, 1989

(6) "Observation of Geographic Update System (GUS) Digitizing Operations in Los Angeles and Denver Regional Census Centers"
Nash J. Monsour to Stanley D. Matchett
November 1, 1989

(7) "Observation of Geographic Update System (GUS) Digitizing Operations in Charlotte and Atlanta Regional Census Centers"
Lawrence R. Ernst to Stanley D. Matchett
January 8, 1990

(8) "Observations of Geographic Update System (GUS) Digitizing Operations in Detroit and Chicago Regional Census Centers"
Robert D. Tortora to Stanley D. Matchett
March 7, 1990

(9) "Observations of Geographic Update System (GUS) Digitizing Operations in Atlanta and Charlotte Regional Census Centers"
Robert D. Tortora to Stanley D. Matchett
March 13, 1990

(10) "Observation of Geographic Update System (GUS) Operations in New York and Philadelphia Regional Census Centers"
Robert D. Tortora to Stanley D. Matchett
April 5, 1990

(11) "Observation of Geographic Update System (GUS) Operations in Seattle and Dallas Regional Census Centers"
Robert D. Tortora to Stanley D. Matchett
April 23, 1990

(12) "Observation of Geographic Update System (GUS) Operations in the Denver Regional Census Center"
Robert D. Tortora to Stanley D. Matchett
May 17, 1990

(13) "Clarifications to Quality Control Program Associated with Geographic Update System Operations"
Robert D. Tortora to Robert W. Marx
June 7, 1990

(14) "Observations of Geographic Update System (GUS) Operations in Boston and New York Regional Census Centers"
Robert D. Tortora to Stanley D. Matchett
July 3, 1990

(15) "Final Observations - Training Sessions on 'Overview of Quality Assurance Programs' for Geographic Update System Operations"
Robert D. Tortora to Stanley D. Matchett
September 4, 1990

SUPPLEMENT:
SUPPORTING MEMORANDA - QUALITY ASSURANCE ASSOCIATED WITH DECENNIAL CENSUS GEOGRAPHIC OPERATIONS

Brian Richards
Statistical Research Division
June 11, 1991

TABLE OF CONTENTS

TABLE OF CONTENTS