DATA GENERATION GUIDELINES

Republic of the Philippines
DEPARTMENT OF LABOR AND EMPLOYMENT
Bureau of Bureau of Labor and Employment Statistics
Intramuros, Manila

October 15, 2004

TABLE OF CONTENTS

Section	**Page**

A. PREPARATORY ACTIVITIES

1. All <u>appropriate</u> tabulations, cross-tabulations or disaggregations of the data categories should have been determined during the Pre-Field Operations stage of the survey. In order that,

    a. tabulations of the same data variables/categories are maximized, hence processing time is minimized;
    b. survey results that would be made accessible to the data users are comprehensive, no need for further tabulations;
    c. consistency checks are facilitated, and
    d. reasonableness of the survey results at detailed levels are easily determined which would not be possible at aggregate levels. For one, errors in coding (industry) may be covered up when the data are tabulated only at higher levels of disaggregation.

2. Based on the proposed data tabulations, the computer syntax should have been prepared prior to table generation.

B. RECONCILIATION OF STATUS CODES AND SELECTED DATA

1. For DUP establishment, track its duplicated establishment or that establishment to be retained.

    a. if the establishment to be retained has responded, it should have the lower EIN. If it does not have the lower EIN, replace it with the lower EIN while the DUP establishment is assigned the higher EIN. Revise name, address, industry and geographic codes if necessary and status of the involved establishments in the status monitoring database.

       <u>Example:</u> Establishment Y with EIN 426 is DUP of Establishment X with EIN 678. Establishment X has responded.

       The EIN of Establishment X should be replaced with 426 and its status code changed to RET. The EIN of Establishment Y should be 678 and its status code becomes DUP of EIN 426.

    b. if establishment to be retained has not responded, it should still have the lower EIN. The necessary revisions in the EIN and other particulars as mentioned above should be made in the status monitoring database for the DUP establishment and the establishment to be retained.

    c. if there are more than one duplicates of an establishment, the principle remains the same. The establishment to be retained has the lower EIN, the duplicates have the higher EINs and the status codes of the involved establishments are adjusted in the status monitoring database.

       ***Notes from Tere:** The problem here is if we change the EINs, it is possible that an EIN may be pertaining to different establishments through time, sometimes the DUP, sometimes the establishment to be retained since we were not keeping close tabs of*

cases like this in previous surveys. As such the <u>establishment history</u> is disrupted. Unless we go back to our previous survey files and check if the integrity of the EIN has been maintained.

What do you think?

2. For CET establishment, track its sample establishments. The status code of these sample establishments should be CON with EIN___ where the EIN should be that of the CET establishment.

CET (consolidated retrieval) is a new code that we have not used in previous surveys. It was introduced in the Office Editing and Data Validation Guidelines of the 2003/2004 BITS and 2004 OWS to pertain to an establishment that provided consolidated or nationwide report. Such establishment used to be coded as RET. However, we have to make a distinction between a CET establishment and those sample establishments whose data are included in the CET establishment. Hence the sample establishments of the CET establishment should now be coded CON with EIN ___. This is done for purposes of assigning the appropriate BUF to the CET establishment.

In case the sample establishment/s of the CET establishment has also responded and its RET status and data values have been encoded, its correct code (CON with EIN ___) should appear in the status monitoring database. Also the data values that have been encoded should be deleted in the respondents' database thus retaining the responses of the CET establishment only.

3. For common samples whose questionnaires were retrieved for 2003/2004 BLES Integrated Survey and 2004 Occupational Wages Survey <u>at the same time or separately</u>, check for similarities in EIN, status code, industry and geographic codes and the responses for data variables common to both surveys, i.e. total employment, time-rate workers on full-time basis (hourly, daily, monthly) and establishment characteristics such as ownership, spread of operations, market, existence of union and collective bargaining agreement pertaining to the same reference period, June 30, 2004.

***Notes from Tere:** EIN and status code have to be checked. It is possible that with adjustments in EIN because of DUP establishments, we end up with common samples with different EIN and status code. Similarities in EIN, status code, industry and geographic codes and total employment are important since we will do the consistency exercise anyway when we update the sampling frame for the next survey round.*

While the OWS does not publish total employment and levels of time-rate workers on full-time basis, the percent distribution of such workers by basic pay and allowance intervals and the median monthly basic pay and median monthly allowance are computed based on the levels. The OWS also estimates the median monthly basic pay and median monthly allowance of establishments according to their characteristics.

or

We only do the consistency checks for EIN, status code, industry and geographic codes and total employment. Consider any data dissimilarities for the other indicators as non-sampling/response errors. Besides employment size coverage of the two surveys differ.

What do you think?

4. Lastly, ensure that the totals of retrieved/processed (RET, CET) and "spoilage" (REF, STR, TCL, CBL, PCL, OSE, OSP, DUP, CON, OTH) questionnaires in the database are the same as their corresponding total

number of questionnaires recorded in FM-BLES 03-3.19 Status Monitoring of Returned Questionnaires (retrieved and "spoilage") and FM-BLES 04-4.7 Monitoring of Data Processing Activities (encoded questionnaires).

Confirm further that the sum of RFV and unaccounted questionnaires is the difference of retrieved/processed and "spoilage" questionnaires from the sample size.

Should there be any discrepancy, it probably would be due to the adjustments on status codes made for DUP establishments and its retained establishment or CET and its CON establishments.

C. REQUIRED STATISTICAL TABLES

1. After adjustments in status codes and reconciliation of selected data have been made, generate the *final survey status report* (FM-BLES 03-3.17 Assessment on the Implementation of Field Operations of BLES Survey/s).

2. To aide in the preparation of the BUFs (Blowing-up Factors), generate the preliminary *Table A - Distribution of Establishments and Weighted Retrieval Rates by Industry Group and Employment Size*. Note that "transfers to and from" of establishments should have been taken into account in this distribution.

   a. For each industry and employment size, ensure that:

   $$\mathbf{N'_{hk,lm} \geq eligible_{hk,lm} \geq n'_{hk,\,lm}} \quad \text{where:}$$

   $N'_{hk,lm}$ — estimated population in the initial stratum k and h and in the post-stratum l and m

   It is estimated based on an eligibility ratio, i.e.,
   $$N'_{hk,lm} = N_{hk} \times eligible_{hk,lm}/n_{hk}$$

   where N_{hk} is population count in the initial stratum k and h and n_{hk} is sample count in the initial stratum k and h.

   ***Notes from Tere:** There is a perceived difficulty here that relates to the CET establishment and its CON samples. Do we adjust the N_{hk}, n_{hk} and $eligible_{hk,lm}$ to consider these establishments before estimating $N'_{hk,lm}$? Do we deduct the CON establishments from the N_{hk}, n_{hk} and $eligible_{hk,lm}$? How do we do it?*

   $eligible_{hk,lm}$ — number of eligible samples (RET, CET, REF, STR, TCL, RFV, CBL, Unaccounted) in the initial stratum k and h and in the post-stratum l and m

   ***Notes from Tere:** We have to decide on the CBL. We used to consider this as ineligible in previous surveys but considered it as eligible in the 2002/2003 BITS. If this is our position does this mean that it will not be taken out of the sampling frame for the next survey round or it will be taken out but was considered as eligible only for purposes of data generation of current survey round?*

$n'_{hk, lm}$ responding samples in the initial stratum k and h and in the post-stratum l and m

b. Evaluate this distribution to determine the stratum (employment size) of the domain (industry) or cell (industry and employment size) that should be collapsed because of low response or non-response relative to the number of eligible samples in the stratum or cell.

If in spite of collapsing, there are still eligible establishments without corresponding responding establishments, a <u>second</u> round of adjustment for non-response should be made. This entails replicating the data values of a similarly situated establishment to the non-responding establishment. Such replicates should be marked in the respondents' database as "Replicate of EIN___."

This is done to ensure that $N'_{hk,lm}$ will be the same as the derived $N'_{hk,lm}$ based on the BUFs.

3. Generate the <u>final</u> *Table A - Distribution of Establishments and Weighted Retrieval Rates by Industry Group and Employment Size.* Note that adjustments for CET and CON establishments, collapsing of strata or cells undertaken, and the second round of adjustments for non-response if any should have been taken into account in this distribution table.

4. Generate the BUFs (ratio of the $N'_{hk,lm}$ to $n'_{hk,lm}$) and attach to relevant establishment record.

   Note from Tere: *We will fine-tune the guidelines on this items 2-3 based on future discussions.*

5. Generate the pre-determined *output tables*.

   Evaluate numerical <u>consistency</u> of the data of a variable in a statistical table and across statistical tables where the same variable appears. In particular, check for consistency across totals of the same data variable that were disaggregated into different categories (industry, employment size, type of workers, type/part/cause/agent of injury, among others).

6. Prepare *publication tables*

   a. Since the presentation of the publication tables may be different from the output tables (e.g. percentages or averages), reference should be made to previously published survey results for comparability with updated survey data. Recent economic developments or issuances, e.g. wage orders between previous and current survey should also be considered to explain any variations in the data.

   b. Part of the validation process also takes into account coherence checks with related survey data generated by BLES or by other establishment surveys (annual labor cost per employee against annual compensation per employee from the Census or Annual Survey of Philippine Business and

Industry of the National Statistics Office; pay for normal regular working time per employee, converted to monthly basis, from BITS against median monthly basic pay from OWS; cost of living allowances and other guaranteed and regularly paid allowances per employee, converted to monthly basis, also from BITS, against median monthly allowance from the OWS; total employment against similar data from the List of Establishments; trend in total employment against that of the employment index released by the National Statistical Coordination Board from the Quarterly Survey of Philippine Business and Industry, among others).

Reference to administrative statistics should also be made to determine coherence of survey data (injury statistics from BITS against those generated from the Work Accident and Illness Report of the Bureau of Working Conditions and employees compensation claims filed at the Social Security System, Government Service Insurance System and Employees' Compensation Commission; BITS-based union and CBA statistics against data sourced from the Bureau of Labor Relations, among others).

Notes from Tere: *The subject matter units should further develop specific coherence guidelines for their statistics. Example: for OWS, there should be some wage rate hierarchy among occupations. The trend in median monthly basic pay should be looked into for comparability with that of the compensation index per employee at nominal terms (from NSCB. For labor cost, the percent share of employers' social security expenditures to total labor cost should be greater than the share of employers' social security contributions to total compensation (from ASPBI) since the labor cost inquiry on this has wider scope. Another case is on the percent share of bonuses and gratuities to total labor cost which approximates 1/12 of total labor cost or that given for the 13th month pay. The trend in total labor cost against that of the compensation index at nominal terms is also another point to look into.*

Coherence does not necessarily mean full numerical consistency. "The coherence of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytical framework and over time."

D. FINAL ASSESSMENT OF ACCURACY IN DATA PROCESSING

Though controls are in place during data collection, and editing/validation and data encoding, there may have been lapses that were overlooked during these stages of survey operation. It is essential then that inaccuracies in data processing, to the extent possible, be finally detected and corrected during output table generation.

Should inconsistencies in output tables be noted, backtracking should start from the validation prooflist list then to the questionnaire itself.

To monitor possible inaccuracies at this point, a standardized form was developed, *FM-BLES 04-4.8 Monitoring of Accuracy in Data Processing*. This instrument together with the other survey monitoring forms will aide survey managers to determine areas for improvement not only in the survey procedures but more importantly in enhancing the skills of data processors.