

## Memo

To: LEHD staff

From: LEHD Senior Management

Date: July 1, 2003 (revised October 15, 2004) (revised December 8, 2004)

### **THIS VERSION IS NOT CONFIDENTIAL**

Re: Disclosure Protocols for Statistical and Research Output

This memo outlines the procedures for statistical and research output disclosure proofing analysis for LEHD products including statistical products (*e.g.*, QWI), research papers and presentations, and intermediate research output.

Note: the tables and memos produced in as a part of the disclosure review process are presumed to be Title 13 confidential. They should never be transmitted via e-mail. Of necessity they are prepared on a Census secure computer system. Leave them there and direct the reviewing person to the location of the files.

#### 1. Statistical Products:

- a. QWI products (or QWI-like products) must use the multiplicative noise methodology.
  - i. Note that business counts are not a disclosure and thus do not require the disclosure methodology described below. However, business counts still require review by an SRF before release.
- b. Products must be reviewed by John Abowd prior to release/distribution with the full set of disclosure analysis described in QWI\_disclosure\_control\_table\_public\_use-20040430.xls. (attached)
- c. Except by prior agreement with John Abowd, data products and related disclosure analysis must be provided to John Abowd at least 1-week prior to release date.

#### 2. Research papers, presentations and intermediate research output:

- a. Standard Census Bureau disclosure rules apply for either person-level or business-level analysis.
- b. For business-level analysis, the following disclosure analysis must be completed:
  - i. For any statistic (*e.g.*, regression coefficient), two types of analysis are potentially required for the sample of businesses used to compute the statistic:
    1. Company counts (not establishment). For example, for a regression statistic on a continuous variable, the number of companies in the estimation sample must be provided.
      - a. Note that for the coefficients of a dummy variable in a regression context, the company counts must be for the

- number of businesses in the cell defined by the dummy variable.
2. The fraction of sales accounted for by the top 1, 2 and 4 companies for the sample underlying the statistic.
    - a. The concentration statistics are not required if the only information on businesses used is employment and payroll.
    - c. For person-level analysis (unit of observation is an individual), person counts for any statistic must be provided.
    - d. These rules apply for any type of statistical analysis including a summary statistics table describing the sample.
    - e. Under these rules, it is difficult to release detailed percentiles of a distribution of the actual data. Two possible approaches:
      - i. Recommended: Estimate a continuous approximation (e.g., kernel density) and compute percentiles from the estimated distribution. For this case, disclosure rules (counts/concentration) for regression analysis apply.
      - ii. Round data sufficiently so that estimated percentiles represent statistic for a sufficiently large number of units.
    - f. Such research output can be reviewed by the program manager or any one of the Senior Research Fellows.
    - g. The program manager or SRF will in turn provide a written paper trail of this review to the files of the program manager and the division director.
    - h. Research must be submitted at least one week prior to release.
    - i. General (obvious but worth emphasizing) guideline: Statistics computed from large samples will require minimal review as person or business counts will be well above thresholds (and concentration of businesses will tend to be low). Statistics computed for detailed industry/geographic areas will require more intensive review.
  3. Special rules apply for Census 2000/ACS tabulations
    - a. A research project is deemed to use Census 2000 data if any variable used in the production of the tables or research results comes from the HCEF/SCEF Decennial Census file system in use at LEHD. (Note: No Census 2000 variables have been integrated into the current production ICF/EHF/ECF file system. All current LEHD uses of Census 2000 data are based on direct extraction from the HCEF/SCEF file system.)
    - b. A research project is deemed to use ACS data if any variable used in the production of the tables or research results comes from the internal ACS files in the LEHD system. (Note: No ACS variables have been integrated into the current production ICF/EHF/ECF file system. GAL modifications for the ACS Place of Work coding system are subject to the special ACS disclosure rules if they are used to prepare summary statistics for release outside of Census.)
    - c. A research project uses a "special tabulation" from Census 2000 or the ACS if it produces a table of results using input files that contain a variable from Census 2000 (definition 3.a) or ACS (definition 3.b). All special tabulations from Census 2000 or ACS must be directly reviewed by the Disclosure

Review Board, except as noted below. See the attached memos for guidelines in preparing such tables. Note, in particular, the population definition rules, the rounding rules, and the required methodology for computing percentiles.

- d. The finest level of detail that may be shown for Group Quarters data is Institutional/ Noninstitutional. There are no exceptions to this rule.
- e. Special tabulations with geographic detail that is national or state-level may be released without prior DRB approval. LEHD disclosure review is still required.
- f. Model-based statistical results (coefficients, standard errors) that were prepared from national or state-level geography may be released without prior DRB approval. If the model includes geographic controls at the sub-state level, the coefficients on these controls may not be released without DRB approval. It is OK to note on the table of coefficients: includes controls for [insert\_geography].
- g. Any research project that envisions publication of results at the sub-state level of geography must have its disclosure control procedures evaluated before the project begins. LEHD uses a different internal Census 2000 file system than the one that is authorized for special tabulations. There is a very good chance that analyses based on sub-state geography from the HCEF/SCEF file system will never be approved for publication in tabular form.