RELEASE OF DATA SUBJECT TO ERROR

OVERVIEW

From the design to the dissemination stage, there are various ways in which errors can enter the statistics. Some errors, resulting form the use of sampling, are random and their extent measurable. Other errors of a non-sampling nature can introduce bias into the results. The NSO has a responsibility to place "quality labels" on all disseminated data in order to assist in their interpretation.

NSO POLICY

In view of the above, it is NSO policy that:

1. Estimates for individual cells of a published table should not be suppressed solely because they are subject to large sampling errors, provided users are adequately cautioned regarding the reliability of the data.

2. Data with serious bias, which may be misleading to users, should not be published.

3. Tables that are not published at a given level of detail, due to the lack of reliability of the estimates in terms of sampling errors, may be released to users with appropriate indications of the sampling errors. Greater caution should be exercised in releasing data that are subject to serious bias.

4. All NSO publications should include statements about data quality and methodology.

5. Wherever possible, information indicating the extent and nature of sampling and non-sampling errors should be provided.

6. Sampling errors may be shown either in the table with the estimates to which they refer, or as a separate table of actual errors, or in illustrative form as a table or graph.

7. In addition to standard errors corresponding to the published estimates, methods should be provided for approximating the standard errors of derivative statistics such as estimates of change, ratios of different estimates for the same period, changes in ratios over a different period, etc. If necessary, supplementary data needed for this purpose, such as correlations, should be presented.

PRACTICAL APPLICATION OF POLICY

♦ Where a considerable number of cells in a table are subject to large sampling errors, it is advisable either to combine some of the classifications or else to refrain from publishing the table.

♦ If, as a result of serious bias, data are likely to be misleading, they should not be published.

♦ If users insist on having data that are subject to high sampling errors, such data may be made available on the understanding that they are not published and that it is the users' responsibility to use the data properly and wisely.

♦ Appropriate cautions, preferably in writing, should always be provided with the release of unreliable data. Users should be requested to present these cautions in any publications, reports etc. in which these data are presented or quoted.

♦ Highly misleading data should never be released to users.

♦ In describing data and its quality, technical terms like sampling error should be explained in as clear and simple a manner as possible.

♦ Any descriptions of error should cover both sampling variability (applicable to sample surveys only) and non-sampling variability (response rates, response bias, non-response, imputation).

♦ Statements on methodology should refer to aspects of survey design which affect the extent of error from these various sources.

♦ Asterisks should be used to indicate those figures that are subject to high sampling and non-sampling errors.

- Users are to be informed of the likely principal sources of error and, where possible, the extent of the error.

Users should be provided with an overview of the data collection and processing methodology in order to enable an assessment of whether the data would satisfy requirements and purpose.