Labour Demand Statistics: Administrative Records

Nur Layali Mohd Ali Khan
Department of Statistics, Malaysia

Dubai, United Arab Emirates

What is administrative data?

by-product of the administrative functions of an agency

often under the authority of a set of laws or regulations, an organization will generally register or record a great deal of information which is needed for the administrative process.

may relate to events or transactions (event based systems) or to people

may be processed and tabulated as statistical outputs

Statistics compiled in this way may be distinguished from data collected directly as part of a statistical census or survey

Administrative data as part of an integrated system of labour statistics

A comprehensive system of labour statistics would include statistics:

from regular and ad hoc household-based surveys and censuses

from regular and ad hoc establishment-based surveys and censuses

as a by-product of administrative systems

Such a comprehensive system would also require:

legislation and mechanisms for coordinating the statistical programme;

national standards for defining concepts and units;

standards for classifying the data collected; and

the regular maintenance of comprehensive population frames for use in selecting samples for household and establishment surveys.

Administrative data as part of an integrated system of labour statistics (cont'd)

Statistics based on the *administrative records* of a ministry of labour (and other agencies) *include those on a wide range of subject-matter topics*, such as conditions of work, employment of non-citizens, unemployment, job vacancies, industrial relations and industrial accidents.

The statistics are based on a *variety of records* relating to *people, businesses, events, transactions*, and so on.

From the statistical point of view, the goal is estimates derived from each administrative system are as compatible as possible with those from household- and establishment-based surveys and censuses.

Labour statistics based on administrative records may be used in:

assessing the productivity and efficiency of an administrative system, as well as providing a quantitative measure of its coverage and in monitoring the performance of the system

assessing and monitoring the economic and manpower situation for use in preparing, evaluating and monitoring action plans,structures and outcomes.

Strengths of administrative records as a source of statistics

Low data collection costs for the statistics-producing agency.

- Since the data has already been collected as part of an administrative function, there are few costs in accessing the data for statistical compilation.

No additional response burden for the respondents.

- Similarly, the units of enquiry (persons, businesses, other organizations) are spared the inconvenience and cost of a separate statistical enquiry. The data which they have already provided as part of an administrative process (registration, application, inspection, notification) may be used for statistical compilation without them having to be involved in a separate statistical enquiry on the same or closely related topics.

It represents a *full count* of the 'clients' of the administrative system:

- A complete count is possible since all the records of the system are available for use in the statistical compilation. This means that statistics can be produced for small groups, such as small areas (districts, towns and provinces), without having to be concerned with problems of sampling precision.

Possible disadvantages of administrative records as a source of statistics

The *types of units* described in the records.

- The units used in some administrative systems may not be the most appropriate to satisfy statistical user needs (for example, jobs versus persons, establishments versus enterprises) or may not use a definition of the unit which is compatible with other statistical sources.
- Some systems register persons or organizations, while others register events (which may occur several times for each individual in a given period).
- Different users may prefer one or the other, but may not be able to extract this information from the system.

Possible disadvantages of administrative records as a source of statistics (cont'd)

The *scope* of the registrations/applications.

- too *narrow* a scope
  - certain categories of units are excluded by design (legal exemptions) or otherwise (illegal non-registration, avoidance),
- too *broad* a scope in that it may include groups which are not of direct interest to a user.
- For example, registered job-seekers at employment exchanges exclude those unemployed who have not registered and may include those employed who are seeking a change of jobs or additional jobs.

Possible disadvantages of administrative records as a source of statistics (cont'd)

The *content* of the data collected.

- The forms used in an administrative system *may not include all the information of interest* for statistical users or for statistical processing.
- The data content may be constrained by legislation, limited resources, or other reasons.
- *Coding of important data may be constrained* because the distinctions needed for administrative use are fewer than what users of the statistical descriptions of the units/events will need.
  - Thus the descriptions used as a basis for coding may be incomplete and the system may not use national standard coding classifications.
- Furthermore, data may be correct at first registration, but *not be subsequently updated*.

Possible disadvantages of administrative records as a source of statistics (cont'd)

The *procedures* for handling data.

- The procedures used in the administrative system are designed to serve the administrative objectives, rules and regulations, and not to provide a basis for valid, reliable and timely statistics.
- Administrative procedures may not require the removal from a data base of expired records, or the updating of job details after a person is first registered.
- The operators of the system are more likely to edit and correct those data which affect their decision making, administrative action or output, and not to give much attention to other data which do not affect their work but which are important for statistical analysis.
- Administrative procedures and the flow of forms through a system may also lead to delays in updating a data base.

Possible disadvantages of administrative records as a source of statistics (cont'd)

High processing costs.

- Since administrative procedures are intended to achieve administrative (rather than statistical) output, considerably more attention may need to be given to detecting and correcting errors, and to coding of information which was not needed for the administrative system but is needed for statistical analysis.
- This processing may require expensive follow up and file amendment.

Evaluating the statistical value of an administrative system

1. Coverage of units.

- What types of units are being recorded (jobs, events, persons, etc)?
- In theory (according to regulations or legislation), who should be included or excluded from the system?
  - This will be determined by the legislation, operating manuals and procedures of the administrating agency.
- In practice, what units do the records cover?
- What incentives are there to encourage registration/reporting to the system? Do many units not register/report?
- Does the agency have sufficient capacity to record all units, or are some omitted due to lack of resources?

2. Range of variables

- What data items are included in the records?
  - What data items are used by system administrators?
  - What items are coded?
  - What are the classification used for coding?
  - What reference periods do the data relate to?

Evaluating the statistical value of an administrative system (cont'd)

3. Frequency

- How often is it possible to extract statistics from the system?
  - This will depend on the type of reporting:
    - (1) Continuous reporting of events (e.g. hirings, separations, accidents) can provide statistics with any frequency;
    - (2) Case-by-case registration (e.g. job seekers, vacancies) can also provide statistics with any frequency;
    - (3) End-of-period reporting (e.g. of income earned, number of employees) can only provide statistics with the same frequency as the reporting.
- Is it possible to extract information at different stages of the administrative process?
  - The system evaluator needs to be aware of the way in which data flows from the time of initial registration to final filing.
    - This includes an analysis of time delays, whether data are added in the process, whether there is any feedback (reverse flow of data), what steps are taken at each stage, how data files are organized and held, and so on.

Evaluating the statistical value of an administrative system (cont'd)

4. Timeliness

- What is the delay between the date of an event, the date of the report, the registration of the report, the processing of the report, the finalization of administrative action on the report, and so on?
  - Statistics compiled from final data sets may be very untimely and it may be useful to consider preliminary statistics based on an earlier stage of the system, if this is possible.

5. Geographic specifications

- Does the system cover the whole country, or only urban or only rural areas?
- Are some provinces excluded?
- Do the registering offices have well-defined geographic catchment areas?
- Do the records include a postal address which will give only an approximate location (e.g. town, district) or do they use a precise geographic reference (e.g. a street address or similar)?

Evaluating the statistical value of an administrative system (cont'd)

6. Validity of variables, definitions and classifications

- Are the variables valid and useful for description and analysis beyond the areas of administrative concern?
- Are the definitions and classifications based on national standards and at least comparable with other data sources?
- The variables recorded in an administrative system are often for one of three main purposes:
  - (1) to determine whether the unit is eligible for a service;
  - (2) to determine the type of services, support or obligation to be provided;
  - (3) to determine the amount of services, support or obligation to be provided.
    - These purposes will determine the definitions and classifications used for the variables in the administrative system.

Evaluating the statistical value of an administrative system (cont'd)

7. Reliability of measurement

- The reliability of the data recorded by the system will depend on the following factors:
  - (1) The incentives for "clients" (those providing data to the system) to give correct or incorrect information;
  - (2) The cost to "clients" of finding the correct information;
  - (3) The probability of being found out if incorrect information has been given;
  - (4) The loss to "clients" if they are found to have given incorrect information;
  - (5) The cost to the agency of controlling the information received;
  - (6) The technical means available to ensure the correct recording of information (e.g. for coding occupation);
  - (7) The gain to the operating agency of correcting wrong information.
- The system evaluation should be aware that the quality of data which are collected but not activelyused or updated by system administrators are likely to be doubtful.

Evaluating the statistical value of an administrative system (cont'd)

8. Consistency over time

- How stable is the administrative system?
- Do regular changes occur in the legal exemptions?
- Are system changes unintentionally introduced as a result of operational changes?
  - For example, do surges in workload lead to unpredictable changes in operational procedures, disruptions in the flow of forms or registrations, or other changes which might affect the comparability of results through time? Changes in the attitude and perceptions of the public will also affect the use and misuse of the system, and hence the quality of the statistics produced by the system.

9. Consistency between local agencies

- Are the same procedures, definitions, scope, etc used exactly in all geographic locations?
- Do different offices apply different procedures depending on their available resources or their use of different operating manuals?

Determining which statistical data to extract

What data items should be used in the statistical compilation?

The key issue here is *the need to balance costs against benefits.*
One must achieve a *compromise* between

(a) the *minimum data set* (and associated resource needs) which will provide sufficient statistics to satisfy basic user needs; and

(b) The *generally larger data* set which users would prefer and which requires additional resources for processing.

Determining which statistical data to extract

Factors to consider:

What is the minimum data requirement to meet basic user needs?

- "nice to know" vs "necessary to know".

What are the costs/resources needed for processing the data items?

- Data items which are costly to verify or edit or difficult to code are less likely to be always completed and may be less reliable.

What data items do system operators use in decision making?

- These are likely to be more carefully checked and therefore more reliable than data items not used by administrators/operators of the system.

What data items are not well completed and therefore less reliable?

- One should examine each item with care and caution. The data may not be as reliable as expected.

Determining which statistical data to extract

The main types of data items which need to be accessed in statistical compilations:

(a) *identification data.*

Each record should have a unique identifier or reference number for ease of filing and retrieving.

(b) *date(s)*

The date on which an event occurred, date of registration or reporting of the event, date on which action was taken, date of death (if different from date of accident), and so on, may all be relevant in order to be able to place a record into a particular reference period for tabulation and for calculating durations;

(c) *Characteristics to classify data*

Each record should have data to be used in grouping or classifying the record when producing tables or graphs. For example, nationality, occupation and age may be used as classificatory data items.

(d) *quantitative data.*

This is the information to be aggregated or averaged. For example, income, age and number of working days lost.

Finally, in selecting data items to be accessed in statistical compilation, one should be aware that the whole process of statistical compilation (including the choice of data items) should be reassessed after an initial period to decide whether costs and quality are appropriate.

Improving the statistical compilation from a system

Existing records may be underutilized.

- Data which were previously unused may be extracted.
- Data which are missing from the record may be available from other sources and linked to the data record for improved statistical analysis

The scope and coverage of the system might be improved by

- amending legislation, introducing (more) incentives and/or disincentives
- providing more resources
- improving procedures.

Improving the statistical compilation from a system (cont'd)

The timeliness of output may be improved

- extracting data from a different stage of the administrative process.

The speed and reliability of reporting and processing may be improved

- redesigning the reporting form
  - A form which is simple and clear in design and layout may reduce the reporting burden and delays in processing, minimize data errors and omissions, and generally improve the quality of statistical outputs. Key requirements include:
    - (1) formulating questions in a clear, unambiguous manner
    - (2) If printed, ensuring that the print is large enough to read
    - (3) providing sufficient space for answers to be recorded in detail.

Improving the statistical compilation from a system (cont'd)

Improved consultation between statisticians, computer system analysts and system administrators.

- These consultations may improve the overall efficiency of the system as well as the operation and design of any supporting computer system, and thereby lead to improved timeliness, content, reliability, etc.

Points to consider

- Before committing to revising or initiating the extraction and compilation of data from an administrative system, it is important to determine the amount and quality of resources available, and what resources are required for the statistical compilation.
  - It may be decided that the cost of these inputs does not justify the statistical outputs achieved.
  - If, however, it is decided to proceed then *care must be taken to adequately plan the project implementation*, including attention to the *timing of inputs, critical paths, approval of budgets*, and so on.

Data processing and output

- Statistical compilation and presentation will usually require data processing and analysis which:
- (a) simplifies details for ease of interpretation (coding);
- (b) checks input records for completeness and accuracy (edit/amendment);
- (c) summarizes data into tables, graphs and charts (tabulation phase);
- (d) interprets and presents the results in a report (reporting).

Data processing and output

Statistical compilation and presentation will usually require data processing and analysis which:

(a) simplifies details for ease of interpretation (coding);

(b) checks input records for completeness and accuracy (edit/amendment);

(c) summarizes data into tables, graphs and charts (tabulation phase);

(d) interprets and presents the results in a report (reporting).

References

Castillo, Monica. Power Point presentation, "Methodologies: Sources of information on labour statistics (focus on the economically active population)", National Labour Market Information Training Programme Port of Spain, Trinidad and Tobago. November 2011

Pember, R. J. (1998). Compilation and presentation of labour statistics based on administrative records. *Bulletin of Labour Statistics*, (1998-1).

End of Session 10