

Error Detection

Quality control procedures to ensure accuracy and quality of data reported are followed by Statistics Canada in cooperation with the Police Information and Statistics Committee (POLIS) of the Canadian Association of Chiefs of Police. Problems with data quality are primarily related to respondent misunderstanding of survey scoring rules and occasional data entry errors. An additional problem that arises with the UCR2 survey is data extraction error.

The data capture and correction processes of both UCR surveys are ongoing activities. At present, the final counts are released to the public in July. This release follows a lengthy data production cycle which begins in December of the previous year with the preparation of interim consolidated statistics. These preliminary statistics are prepared to provide an early warning of any data problems.

Problems are identified with the use of a statistical “tolerance” model specifically designed to compare a single respondent to its past behaviour and to that of similar respondents. Problems are further identified by an “outlier detection” model which creates a number of ratios internal to a particular police service and offence. Based on the significance of the outlier, it is flagged for further investigation.

Non-respondents and data anomalies are easily identified using these models, and staff then follow up with the police service.

The larger police departments responding to the UCR2 surveys are asked to review summary data tables for completeness and accuracy and, if necessary, adjustments are made before a master file is created. Each of these police services is required to “sign off” on their data prior to release.

In order to create either the preliminary or final master file, the incident-based file must be converted to produce counts comparable to those produced by the aggregate survey. The converted counts, or the “Derived Aggregate Offence” data, are combined with the aggregate survey data to produce national and provincial tables. In this way, continuity is maintained with the historical crime series.

Data quality is of primary concern with respect to any survey. The CCJS follows several different practices to ensure the quality of UCR data. The internal integrity of records received from respondents and the reasonableness of a respondent’s counts are continually assessed through both manual and electronic editing of the data. Where possible, UCR data are checked against independently produced respondent crime statistics such as police annual reports.

The CCJS does not, however, normally audit respondent data. As the aggregate and incident-based surveys operate in a significantly different manner, specific issues surrounding data quality and the procedures required to ensure it are different.

Aggregate UCR Survey

Compared to the Incident-based UCR Survey, the aggregate survey is fairly simple to operate, validate, edit and release. The aggregate master files are continually updated throughout the year, and are further processed when necessary to produce consolidated aggregate statistics. Errors identified during the editing of the data are pursued with the respondent where necessary and corrected. As the data reported to the aggregate survey are provided in the form of summary counts, problems related to counting and scoring can only be identified by assessing the reasonableness of a respondent’s counts. In this regard, extensive survey support to respondents through user guides, telephone support and training is crucial to maintaining data quality.

Incident-based UCR2 Survey

The Incident-based UCR Survey captures a broad variety of data on each criminal incident. Data from this survey also undergo extensive editing during the data capture process at the CCJS. Unlike aggregate data where editing can only identify missing data and ensure that cell values tabulate correctly, incident-based data are subjected to very detailed validity and logic checks and imputations which assess a respondent's compliance with survey data requirements and ensure that only records which conform to these requirements are available for analysis.

Consequently, in addition to being able to assess the overall counts of incidents, the quality of data associated with each incident can be evaluated in greater detail. Anomalies in crime reporting practices that previously lay undetected in aggregate data are now being corrected. As more police agencies begin to participate in the Incident based UCR2 Survey, the data provide an increasingly more accurate picture of the nature of incidents reported to police in Canada.

Specific areas of concern with respect to data quality include: the variety of software interfaces used to extract incident-based UCR2 data; the complexity of data extraction and data capture; the different operational and administrative practices among different police forces, and; the need to accurately translate these data into aggregate counts.