General Social Survey
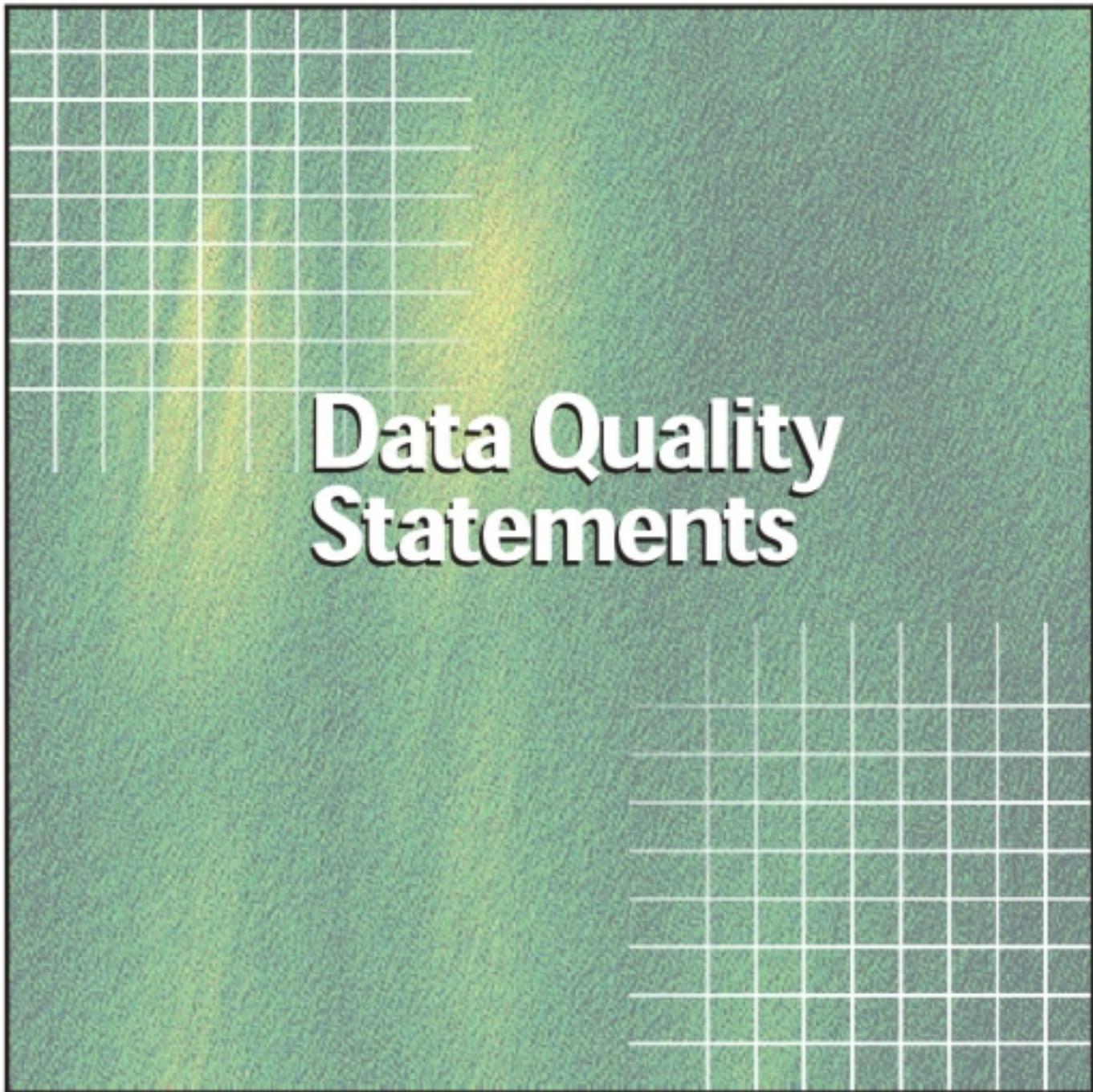
Statistical Data Documentation System
Reference Number 3895, 3897, 4500, 4501

Data Quality
Statements

Statistics Canada Statistique Canada

Canada

General Social Survey

Statistics from the General Social Survey (GSS) databases are estimates based on data collected from a small fraction of the population (roughly one person in 2,000) and are subject to error. The error can be divided into two components: sampling error and non-sampling error.

Sampling error is the difference between the estimate derived from a sample and the result that would have been obtained from a population census using the same data collection procedures. For a sample survey such as the GSS, this error is estimated from the survey data. The measurement of error used is the standard deviation of the estimate. When a sampling error is more than 33 1/3% of the estimate itself, it is considered to be too unreliable to be published. In such a case, the symbol " -- " appears in statistical tables in place of the estimate. When the sampling error is between 16 2/3% and 33 1/3%, the corresponding estimate is accompanied by the symbol " * ' in a table. Such estimates should be used with caution. Finally, all estimates with a sampling error of less than 16 2/3% can be used without restriction.

All other types of errors, such as coverage, response, processing, and non-response, are non-sampling errors.

Many of these errors are difficult to identify and quantify.

Coverage errors arise when there are differences between the target population and the surveyed population. Households without telephones represent a part of the target population that was excluded from the surveyed population. To the extent that this excluded population differs from the rest of the target population, the estimates will be biased. Since these exclusions are small, one would expect the biases introduced to be small. However, since there are correlations between a number of questions asked on this survey and the groups excluded, the biases may be more significant than the small size of the groups would suggest.

Individuals residing in institutions were excluded from the surveyed population. The effect of this exclusion is greatest for people aged 65 and over, for whom it approaches 9%.

In a similar way, to the extent that the non-responding households and persons differ from the rest of the sample, the estimates will be biased. The overall response rate for the GSS was approximately 80%.

Non-response could occur at several stages in this survey. There were two stages of information collection: at the household level and at the individual level. Non-response at the household level averaged 6%. Non-response also occurs at the level of individual questions. For most questions, the response rate was high and, in tables, the non-responses generally appear under the heading "not stated".

While refusal to answer specific questions was very low, accuracy of recall and ability to answer some questions completely can be expected to affect some of the results presented in the subsequent chapters. Awareness of exact question wording will help the reader interpret the survey results.

Since the survey is cross-sectional, caution is required in making causal inferences about the association between variables. Observed associations may be a reflection of differences between cohorts, period effects, differences between age groups or a combination of these factors.