Data quality

The methodology of this program has been designed to control errors and to reduce the potential effects of these. The food statistics program is collecting data from administrative sources that originate from existing survey's data sets and from other various sources. The sampling errors from each survey have been controlled and minimized by the team responsible for the survey. To know more on sampling errors of a particular survey, the user must refer to the methodology of the particular survey. The food statistics program is not revising the sampling errors from the surveys. In case of a situation where the data from a survey are questionable, the analyst of the food statistic program will inform the section responsible for the survey of the finding. If data need to be revised, it will be done by the section responsible of the survey. Then after, the corrections will be incorporated to the food statistics program. The other possible source of errors is non-sampling errors. Examples of non-sampling error are data response error, non-response error and processing error. A discussion of these types of errors and the steps taken to address them follows.

Administrative data that originates from sources other than surveys may contain data response errors. They may be due to inability or unwillingness of the respondent to provide correct information, misinterpretation of the questions or definitional problems. For the food statistics program, these errors are controlled through the use of simple concepts, data confrontation and consistency checks.

Non-response error is related to respondents that may refuse to answer, are unable to respond or are too late in reporting. For the survey, this type of error is mitigated by the close contact Statistics Canada staff maintains with respondents.

Processing error may occur at various stages of processing such as data entry, editing and tabulation. For the program, various measures have been taken to minimize these errors. For instance, data entry and edit are performed simultaneously due to the spreadsheet design which allows errors to be quickly seen. As well, historical ratios aid in eliminating outliers created by data entry. Tabulation is automated to eliminate human error. Finally, multilevel checking is performed to detect any errors.

Data Limitations

All statistical programs have limitations and readers should keep in mind that some factors can be both strength and a limitation. This is true with the Canadian supply based food and nutrition data. One of the limitations of the data set is that it is apparent. Food available or food available adjusted for losses are derived and not measured directly because the cost and burden of measuring consumption directly on an ongoing basis would be overwhelming. As a residual calculation any errors in the input data that are not compensated for will flow to the residually derived food available estimate. As this number is the basis for all further calculations, any error would then flow to the food available and nutrient data.

The program relies on a quantity of quality inputs. Sometimes, the inputs are not available, are discontinued or are suspect. This requires constant analysis and maintenance to sustain the quality of the overall program. Further, some commodities are estimated based on related information as opposed to being derived directly. The program does have some data gaps that often refer to less common commodities or less relevant data inputs. Also, some sources of nutrients such as supplements are not accounted for.

The food statistics program is national in scope and is therefore unable to provide geographical breakdowns at the provincial or sub-provincial level. The static model relies on a large number of factors such as nutrient equivalent factors or waste adjustment factors. Any error in a factor would be carried to the output data unless there are compensating errors. The nutrient data utilise factors that are held constant over time. If the nutrient values of foods changes importantly with time, this characteristic is not accounted for. The loss adjustment factors are provided by the United States Department of Agriculture's Economic Research Service. Cultural or even climatic differences might affect these numbers, however, they are provided as estimates and considerable caution needs to be used when working with the loss adjusted data. Losses, particularly losses at the consumer level, are by their nature difficult to estimate. The losses are held constant over time but the loss rates may vary. All of the nutrient factors and loss adjustment factors are based on weight. Consequently, the loss adjustments have a limitation in that they may only partially account for losses of some nutrients as a result of cooking. For instance, more (or less) fat, fat soluble nutrients or water soluble nutrients may be lost depending on the methods used in preparation and cooking.

Although every attempt has been made to adjust the nutrient data to approximate consumption, the data are estimates, a limitation that readers should keep in mind when working with the files. Further, these data do not account for nutrients that may be provided by dietary supplements or from the water supply. Except in the case of sugar containing products and some meat based products, the export data do not account for the import or export of processed products that contain various types of food. If the export of food containing products is considerably different than the import of food containing products then the residually calculated net supply and food available will be biased. Although there are numerous commodities estimated, often these commodities are consumed indirectly as ingredients in processed foods. For instance, wheat flour is accounted for while bread is not estimated directly. The supply side data have a limitation in that they cannot be used for certain types of appropriate and important analysis. The data cannot be broken down by age, sex, geography, weight, health or other characteristics that are relevant to the analysis of food statistics.

Key descriptions that might be used to portray the program include:

• relevant

- by commodity -allows agriculture or health related analysis
- time series -allows trend analysis
- inexpensive, efficient, adds value
- credible basis for analysis
- national in scope
- apparent –relies on model (consistent factors)
- in general, credible input data from numerous reliable sources
- national and international cooperation
- concerns related to the quality of some inputs or data gaps
- USDA based waste factors.