Statistics Canada Statistique Canada

Challenging the Data: Job Aid

This job aid is designed to help you critically assess the data presented to you. No data is perfect. By understanding the strengths and limitations of the data, you can avoid being misled—and make smarter, more informed decisions.

Source Can you identify the specific data source? If the estimate comes from a report, can you identify its original source? Is the source reliable and trustworthy? Are there known limitations or caveats associated with the data used to produce the estimate?
Methods How was the data obtained—was the methodology transparent? Did it come from a sample survey of a well-defined population of interest, a marketing poll, an administrative database, or did it reflect insights from a qualitative study like a focus group?
Timeliness Is this data current? When was this dataset acquired? Do you know when the data was last updated? Does timing affect its validity?
Sample What is the sample size? What is the sampled unit (e.g., individuals, households, businesses)? What is the response rate (out of all sampled units, what portion responded)? How were respondents chosen (were they randomly selected, or did they volunteer)? Does the sample represent the population of interest (for example, all Canadians) or are some units systematically excluded from the sample, resulting in poor coverage of this population?
Definitions and measures How is the concept of interest defined, and how is it being measured? Is the definition based on a recognized standard[1]? If you are comparing estimates from different sources, were the underlying concept definitions and measurements the same, or at least comparable?
Accuracy How was the estimate calculated? Were quality[2] indicators made available to assess the estimate's fitness for use?

Canada

Context

What is the broader context of the data?

How do these estimates compare…

… to estimates from other data sources?

… to similar estimates from previous years, if available?

… across groups based on current knowledge about group differences?

Has sufficient background been provided to acknowledge the lived experience or historical realities of the population(s) to which the data refers?

Disaggregation

Are the data available at the right level of disaggregation[3] for your analysis?

Additional data sources

Are there other data sources that could provide a more nuanced understanding of the concept being measured?

Could linking[4] multiple datasets reveal new information?

Ethical considerations[5]

What is the expected public benefit of using this data?

What are the potential negative consequences of using this data (e.g., privacy intrusiveness, lack of transparency, harm to individuals or groups, trust erosion, sustainability issues, security concerns, etc.)?

Are the potential benefits credible considering the potential consequences?

---

[1] More information on standards used at Statistics Canada can be found on Standards, data sources and methods (https://www.statcan.gc.ca/en/concepts/index).

[2] Quality indicators should describe precision (e.g., confidence intervals) as well as potential bias that could arise due to low response rates (if respondents are different from non-respondents) or poor representativeness / coverage of the population of interest in the sample data. For an introduction to the fundamentals of data quality, watch the video *Data quality in six dimensions* (https://www.statcan.gc.ca/en/wtc/data-literacy/catalogue/892000062020001). For more information on confidence intervals, consult *Understanding Confidence Intervals for the 2021 Census* (https://www12.statcan.gc.ca/census-recensement/2021/ref/98-20-0001/982000012021003-eng.cfm).

[3] Disaggregated data is data that is broken down at the lowest level of detail possible. Disaggregating data helps uncover differences that the big picture can hide. Consult Statistics Canada's disaggregated data standards (https://www.statcan.gc.ca/en/concepts/dds).

[4] Statistics Canada has the infrastructure to support data linkages across multiple domains, such as health, justice, education, and income. More information can be found on:

- Social Data Linkage Environment (SDLE) (https://www.statcan.gc.ca/en/sdle/index)
- Business - Linkable File Environment (https://www.statcan.gc.ca/en/about/statcan/lfe)

[5] For more information on data ethics, watch the video *Data ethics: An introduction* (https://www.statcan.gc.ca/en/wtc/data-literacy/catalogue/892000062022001). You can also consult this report *The Foundations of ethical reviews at Statistics Canada* (https://www150.statcan.gc.ca/n1/pub/11-633-x/11-633-x2024002-eng.htm).