DATASSIST
Centre for Social Innovation
44-6 Wellesley Place
Toronto, Ontario M4Y 3E1
905-609-5483
heather@idatassist.com
www.idatassist.com

Checklist for Data Integrity

Did you:

- ☐ Document all the notes and findings from your data audit process?
- ☐ Write out the specific research questions and hypotheses—to remind you of what the data is being used for?
- ☐ Check with the data collection team to see if they are aware of any data collection issues?
- ☐ Spot-check the data? – it should be proofread against the original surveys to see that it has been entered correctly.
- ☐ Check the indicators one at a time?
  - Check to see that the data for this variable contains only possible values. (For example, nothing bigger than the maximum possible value.)
  - Check whether the data contains unlikely values. (For example, weights which seem unlikely.) *Flag these values.*
- ☐ Check the indicators two at a time?
  - Create a few tables with the data to check that the data is possible across multiple indicators. (Are there unemployed with employment income? Are there young children with dependents?)
- ☐ Plot the data?
  - Plot the data in many different ways – the ways you personally like the best and simply look at it. Plot the variables by themselves. Plot 2 and 3 variables together.
- ☐ Compare the current data with any past data that is available from the survey?
  - Are they similar enough to make sense?
- ☐ Compare the current data to any available secondary data?
  - Are they similar enough to make sense?
- ☐ Deal with missing data?
  - Calculate how much of each indicator is missing.
  - If most data is not missing, simply ignore the missing data.
  - If there is significant missing data, consult dealing with missing data guides and make a decision.
- ☐ Create any derived variables you need to answer your research question?
  - Recode and derive variables that may be necessary to properly answer the questions.
  - For example, calculate BMI from height and weight indicators.