

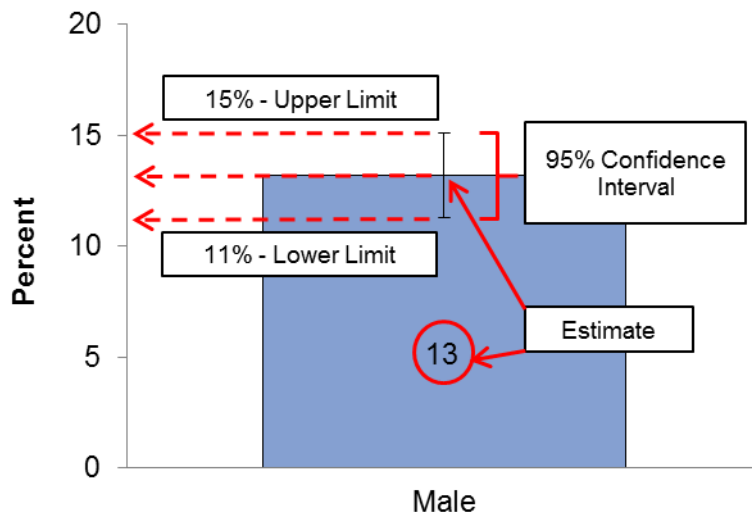
Halton Health Statistics Data Interpretation

This guide to interpreting data will provide information on common statistical tools and terms used in reports produced by the Halton Region Health Department. Statistics and graphs are commonly used to summarize, describe and gain a better understanding of public health data.

Confidence Intervals:

In public health, estimates are often presented in the form of percentages with confidence intervals. A **95% confidence interval** refers to the range of values that has a 95% chance of including the true estimate. Confidence intervals are reported in brackets next to the estimate in the text with a “±” symbol (e.g. 13% (±2)) or presented as “I” shaped bars on graphs.

In the example in **Figure 1**, using hypothetical data, 13% (±2) of adults aged 18 and over in Halton Region are current smokers. As we have not asked 100% of the Halton region population if they smoke, this percentage is an estimate of the true percent of current smokers in Halton. If we re-surveyed a different sample of people in Halton region, the estimated percent of current smokers may differ slightly because different individuals would be included in the sample.



In the example in **Figure 1**, if we were to repeatedly survey people in Halton region about their smoking status, 95 times out of 100 we would expect the percent of current smokers to fall within the 95% confidence interval—in this case between 11% and 15%.

A 95% confidence interval can be used to describe the **reliability** of an estimate. An estimate with a narrow confidence interval, as in the 45-64 age group in **Figure 2**, indicates that the true value is likely very close to our estimate. An estimate with a wide confidence interval, such as the confidence interval for ages 18-24 in **Figure 2**, suggests that our estimate could be further away from the true value.

The **reliability** of an estimate can be influenced by sample size. Small sample sizes tend to result in less reliable estimates.

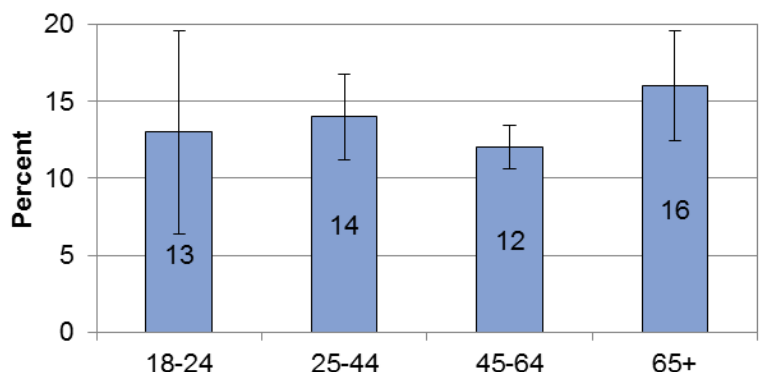


Figure 2: Example of confidence intervals with differing precision

Statistical Significance:

When comparing two estimates to one another, we refer to any difference between the two estimates as being either statistically significant, or not statistically significant. A **statistically significant** difference between two estimates is a difference that is likely not due to chance alone.

Confidence intervals can be used to compare two estimates to determine if the difference between the estimates is **statistically significant**. If the confidence intervals do not overlap when comparing two estimates, the differences are statistically significant and likely not due to chance (see **Figure 3**). If the confidence intervals do overlap when comparing two estimates, the differences may be due to chance and are not statistically significant (see **Figure 4**).

Since overlapping confidence intervals are used to determine statistical significance, p-values are not calculated for Halton health statistics reports. This is a conservative approach ($\alpha < 0.01$) which is more appropriate when multiple comparisons are being made, as in many of the Halton Health Statistics reports.

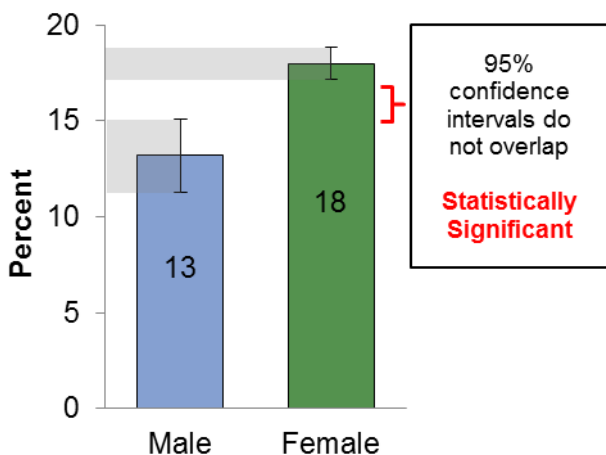


Figure 3: Difference between males and females is statistically significant

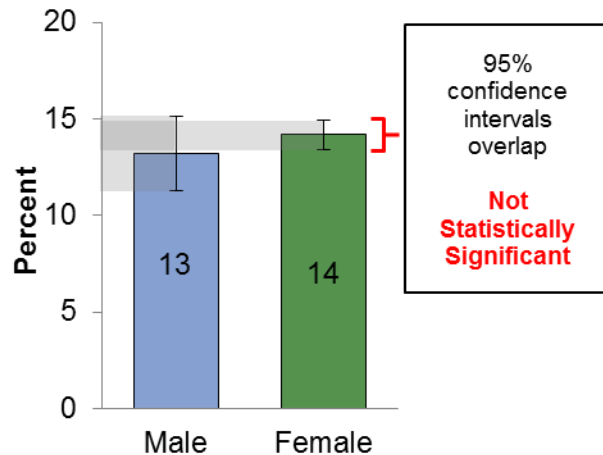


Figure 4: Difference between males and females is not statistically significant

Coefficient of Variation:

The **coefficient of variation** measures the amount of variability of data points around an estimate. Like the **confidence interval**, the coefficient of variation indicates how reliable or precise an estimate is. Estimates with a larger confidence interval have a larger coefficient of variation. **Figure 5** shows an example of a graph with data that has a high degree of variability.

A **coefficient of variation** between 16.6 and 33.3 indicates a high amount of variability. These estimates may not accurately reflect the true population trends and should be interpreted with caution; they are marked with an asterisk (*) in the graphs and tables. Estimates with a **coefficient of variation** of 33.3 or greater are not reportable because they likely do not accurately reflect the true population trends, and are marked with double asterisks (**) in the graphs and tables.

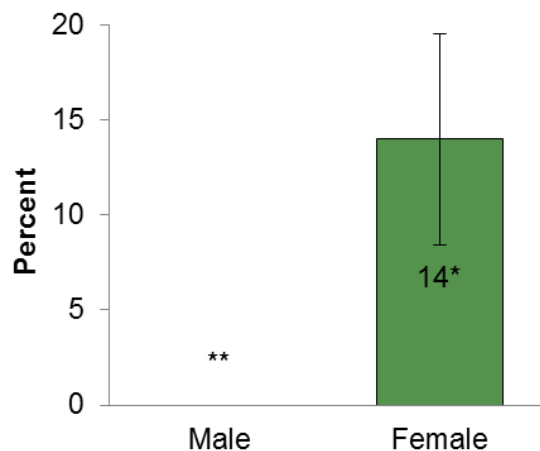


Figure 5: Example graph of data with high degree of variability

Adapted from: KFLA Public Health. (n.d.). Understanding Confidence Intervals and Statistical Significance in Facts & Figures.