**Equations for clustered diagnostic accuracy studies**

Adapted from:

Kwak, M., Um, S.-W., & Jung, S.-H. (2015). Comparison of operational characteristics for binary tests with clustered data. *Statistics in Medicine*, *34*(15), 2325–2333. https://doi.org/10.1002/sim.6485

**Sensitivity:**

An unbiased estimator of the sensitivity for a diagnostic test p, is:

Let denote the total number of diseased sites across all n subjects. We then have:

Because the n subjects are independent, given the disease status of the patients,

are independent 0-mean random variables. By the central limit theorem for large n,

is asymptotically normal with mean 0 and variance that can be consistently estimated by

**Specificity:**

An unbiased estimator of the specificity for a diagnostic test ,p, is given by:

Let denote the total number of non-diseased sites across all n subjects. We then have:

Because the n subjects are independent, given the disease status of the patients,

are independent 0-mean random variables. By the central limit theorem for large n,

is asymptotically normal with mean 0 and variance that can be consistently estimated by

**PPV:**

A consistent estimator of PPV is:

Define N as the total number of sites across all n subjects:

If we define:

Then:

The ignorable term, , is added by replacing the consistent estimator    with in the denominator. This gives us:

Where for are independent random variables with mean 0. Hence, is approximately normal with mean 0 and variance that can be consistently estimated by

Where

**NPV:**

A consistent estimator of NPV is:

If we define:

Then:

The ignorable term, , is added by replacing the consistent estimator    with in the denominator. This gives us:

Where for are independent random variables with mean 0. Hence, is approximately normal with mean 0 and variance that can be consistently estimated by

Where