Appendix 3. Summary of formulas used for the calculation of bias and effective sample size.

Summarized from Bross I. Misclassification in 2 x 2 tables. International Biometric Society 1954;10:478-486.

Bias in Prevalence Estimate

Let:
n= number of individuals drawn from population
u= actual number of "problem" cases in sample (true positives=people with HAND, classified as having HAND)
v= number of "problem" individuals incorrectly classified as "non-problem" cases (false negatives=people with HAND, classified as no HAND)
w= number of "non-problem" individuals who are incorrectly classified as "problem" cases (false positives)
x= apparent number of problem individuals

Therefore:
x=u-v+w

Let:
p = proportion of "problem" cases in the population; true prevalence
q = is the proportion of "non-problem" cases
θ = probability of misclassifying a "problem" individual, e.g. false negatives
ϕ = probability of misclassifying a "non-problem" individual, e.g. false positives

If misclassification exists, the estimated proportion of "problem" cases will not be p as above but instead:
$E(x/n) = p - p\theta + q\phi$ where $q = (1-p)$

Then:
$E(x) = n\ p\ K$ where $K=1- \theta + q/p\ (\phi)$

Thus:
Estimated prevalence = $E(x/n) = p\ K$

And:
Bias in prevalence estimate = $p\ K - p$

Effective sample size

The quantity K' measures the effect of misclassification on the variance of the estimate p where:

$K' = 1-2\phi - 2\theta + \theta/q + \phi/p$

The accepted measure of efficiency is:

Efficiency = $1/(K'+D)$ where $D= [(q\phi - p\theta)^2\ n]/pq$

And an accepted measure of effective sample size is:

Effective sample size= n * efficiency