# MedicalBiostatistics.com

## Measures of Morbidity in a Community

# Capture-Recapture Methodology



Adapted from Medical Biostatistics, Third Edition (MedicalBiostatistics.synthasite.com)
by A. Indrayan (indrayan.weebly.com)
Chapman & Hall/ CRC Press, 2012 US$129.95
Available at CRC Press

This method was originally devised for obtaining counts of animals in the wild. Suppose a sample of $n_1$= 48 deer is captured, marked, and then released. They mix with other deer. Subsequently, a second sample of $n_2$ = 35 deer is captured from the same area. If $m$ = 16 of them are found to have the mark, an estimate of the total number of deer is obtained by inflating by a factor of (35/16). Thus, an estimate of the total count $N$ is 48×35/16 = 105. Statistically better estimate in the long run is obtained when one is added to each of these numbers and finally one is subtracted. Thus, the

$$\text{Capture-recapture estimate: } N = \frac{(n_1+1)(n_2+1)}{m+1} - 1, \qquad (1)$$

where $n_1$ is the size of the first sample, $n_2$ of the second sample, and $m$ is the number of cases found to appear in both samples.

In health and medicine, the capture-recapture methodology helps to estimate the total number of cases when their incomplete count is available from two *independent* sources. These sources could be hospital records, physicians in private practice, death certificates, or any other such list of cases. The count of duplicate cases, which appear in both lists, can help to substantially improve the estimate of the total cases. Hook and Regal argued that capture-recapture method can improve prevalence estimates even for apparently exhaustive surveys.

**Table 1** Capture-recapture estimate of childhood diabetes in Madrid

| | Hospital records | Diabetes Association | Common | Capture recapture estimate |
|---|---|---|---|---|
| Number of cases | 432 | 138 | 119 | 501 |
| evalence rate per 100,000 population (total population 4.4 million) | 9.8 | 3.1 | | 11.4 |
| Extent of undercount (as percent of the last column) | 14% | 72% | | 0% |

**Example Capture-recapture estimate of childhood diabetes cases**

The methodology was adopted to assess the prevalence of childhood diabetes in Madrid, Spain. A population-based registry identified 432 cases through hospital inpatient records. Another source was the Spanish Diabetes Association, which recorded 138 cases. It was found on matching that 119 cases were common to the two sources. Thus, an estimate of the total number of cases of childhood diabetes using Eq. (1),

$$= \frac{(432 + 1)(138 + 1)}{119 + 1} - 1$$

$$= 501.$$

The results can be converted to prevalence rates as shown in Table 1.

The cases in either source are undercounts but the duplicates help to come up with an estimate of the total cases as well as an improved prevalence rate.

Note the following for capture-recapture method.

1. The capture-recapture methodology assumes that there are no intermediary additions or deletions. If there are, the estimate may have to be revised accordingly.
2. It is necessary that the 'captured' subjects move around freely and are homogeneously mixed when 'released'.
3. The two sources should be independent of one another. This is not fulfilled in a situation where, for example, some cases seen by the first source tend to be referred to the second. In such cases, the capture-recapture estimate may be too low.
4. Like all estimates, the estimate of the total count by capture-recapture methodology is subject to sampling fluctuation. Methods are available to find the standard error of this estimate, and to construct a 95 percent confidence interval.