The unique feature of the dual system estimator is that it adjusts for population underenumeration due to incomplete reporting. It is based on the estimate derived from each of two imperfect data collection systems and on an estimate of the number of the same individuals that are counted by both imperfect systems. The dual system estimator of \( N \), the population size, is:

\[
\hat{N} = \frac{\hat{N}_1 \hat{N}_2}{\hat{N}_{12}}
\]

where \( \hat{N}_1 \) is the estimate of \( N \) based on the first data collection system, \( \hat{N}_2 \) is the estimate based on the second system, and \( \hat{N}_{12} \) is the estimate of the number of individuals that are counted by \( \hat{N}_1 \) and \( \hat{N}_2 \).

Even though \( \hat{N}_1 \) and \( \hat{N}_2 \) are biased estimates of \( N \), the dual system estimator of \( N \) is a consistent estimate of \( N \) if the following assumptions hold:

1. Every individual in the population is eligible to be enumerated by one or both data collection systems.
2. Whether the individual eligible to be enumerated by a system is reported or not is a Bernoulli variable.
3. Whether the individual, eligible to be enumerated by both systems, is reported or not reported by one system is independent of whether or not he is reported by the other system.

The traditional dual system estimator assumes that \( \hat{N}_1 \) and \( \hat{N}_2 \) are based on different collection systems, such as a household sample survey and a registration system, and that \( \hat{N}_{12} \) is determined by matching the files of the individuals enumerated by the two systems. The dual system network estimator is based on procedures that are quite different.

The dual system network estimator assumes that \( \hat{N}_1 \) and \( \hat{N}_2 \) respectively are based on disjoint counting rules \( r_1 \) and \( r_2 \) adopted collaterally by a household survey, henceforth referred to as a dual system survey. \( \hat{N}_{12} \) is based on a quality check survey using the following scheme that was proposed by Nathan [2]. Quality check interviews are conducted at households that are eligible to report individuals by a different counting rule than the one by which they were enumerated originally in the dual system survey. Thus, if an individual was reported in compliance with \( r_1 \) in the dual system survey, he would be enumerated at a quality check household eligible to report him in compliance with \( r_2 \) and vice versa. The objective of the quality check is to ascertain whether or not the individual previously reported in the dual system survey is also reported by another household that is eligible to report him by the other rule.

For instance, \( r_1 \) could denote the de jure residence rule that counts individuals at their usual places of residence, and \( r_2 \) a family network rule that counts individuals at residences of their close relatives. Individuals that had been reported by their usual places of residence in the dual system survey, would be enumerated at the households of close relatives in the quality check survey, and the individuals that had been reported by relatives’ households in the dual system survey would be enumerated at their own households in the quality check survey. The quality check households are identified by respondents in the dual system survey. Thus, the individuals that report themselves in the dual system survey provide the names and addresses of relatives that would be eligible to report them in the quality check survey, and the relatives who report individuals in the dual system survey provide the names and addresses of these individuals for follow-up in the quality check survey.

The dual system network estimator and its component estimators are derived in the following sections of this paper. Biases of the traditional dual system estimator and of the dual system network estimator are briefly discussed in the final section.

**ESTIMATORS \( \hat{N}_1 \) AND \( \hat{N}_2 \)**

The \( I_a \) \((a=1, \ldots, N)\) individuals in a population are enumerated in the dual system household survey by two mutually exclusive counting rules, \( r_1 \) and \( r_2 \). The links between \( I_a \) \((a=1, \ldots, N)\) and the households \( H_i \) \((i=1, \ldots, M)\) eligible to report them in compliance with these rules are specified by

\[
\delta_{ai} = \delta_{ai1} + \delta_{ai2}
\]

where the indicator variable

\[
\delta_{aik} = \begin{cases} 
1 & \text{if } I_a \text{ is linked to } H_i \text{ by } r_k \ (k=1,2) \\
0 & \text{otherwise.}
\end{cases}
\]

Let

\[
N_k = \sum_{a=1}^{N} \sum_{i=1}^{M} \delta_{ai1} s_{ak} + \sum_{i=1}^{M} \delta_{ai2} s_{ak}
\]

the number of the \( I_a \) \((a=1, \ldots, N)\) eligible to be enumerated by \( r_k \) \((k=1,2)\) and

\[
N_{12} = \sum_{a=1}^{N} \sum_{i=1}^{M} \delta_{ai1} s_{ak} + \sum_{j=1}^{M} \delta_{aj2} s_{aj}
\]

the number of the \( I_a \) \((a=1, \ldots, N)\) eligible to be enumerated by \( r_1 \) and \( r_2 \), where

\[
s_{ak} = \sum_{i=1}^{M} \delta_{aik} = \text{the number of households eligible to report } I_a \text{ by } r_k \ (k=1,2).
\]

If \( s_{ak} \geq 1 \ (a=1, \ldots, N) \ (k=1,2) \) and \( N_k = N_{12} = N \). On
The other hand, if \( s_{a1} \geq 1 \) \((a=1,\ldots,N)\) but \( s_{a2} = 0 \) for some \( I_a \), \( N_1 = N \) but \( N_2 = N_{12} \neq N \). And if \( s_{a1} = 0 \) and \( s_{a2} = 0 \) for some \( I_a \), \( N_1 \neq N_{12} \) and \( N_2 \neq N_{12} \) unless the identical \( I_a \) are covered by \( r_1 \) and \( r_2 \).

The \( I_a \) \((a=1,\ldots,N)\) may or may not be reported in the survey by the \( H_i \) \((i=1,\ldots,M)\) to whom they are linked by \( r_1 \) and \( r_2 \). Consequently, we denote the survey responses by the random variables

\[ \delta_{aik} = \begin{cases} 1 & \text{if } I_a \text{ is linked to and enumerated at } H_i \text{ by } r_k \text{ (}k=1,2\text{)} \text{ and } N_1 = N_2 = N_{12} = N \text{,} \\ 0 & \text{otherwise.} \end{cases} \]

We assume that the \( E(\delta_{aik}) = P_k \delta_{aik} \) \((a=1,\ldots,N)\) \((i=1,\ldots,M)\) \((k=1,2)\), where \( 0 < P_k \leq 1 \).

A random sample of \( m \) out of \( M \) households is selected, and the \( I_a \) \((a=1,\ldots,N)\) at the \( m \) households are enumerated by \( r_1 \) and \( r_2 \) in the dual survey. Two estimates of \( N \) are derived from the survey.

\[ \hat{N}_{i} = \frac{M}{m} \sum_{i=1}^{M} \alpha_i \sum_{a=1}^{N} \alpha_{aik}/s_{aik} \text{ (}k=1,2\text{)} \]

where the Bernoulli variable

\[ a_i = \begin{cases} 1 & \text{if } H_i \text{ is selected in the dual system survey} \\ 0 & \text{otherwise.} \end{cases} \]

Since \( E(\hat{N}_{i}) = \hat{N}_i P_k \delta_{aik} \) \((k=1,2)\), it follows that \( \hat{N}_i \) is an unbiased estimate of \( N \) if and only if \( s_{aik} \geq 1 \) \((a=1,\ldots,N)\) and \( P_k = 1 \).

**ESTIMATOR \( \hat{N}_{i2} \)**

Subsamples of reports of individuals enumerated by \( r_1 \) and \( r_2 \) in the dual system survey are selected for the quality check survey. For each type of report, a quality check household is selected at random from \( s_{at} \) households \((t=3-k)\), where the \( \hat{I}_a \) \((a=1,\ldots,N)\) is enumerable by a different rule than the one by which he was enumerated originally. Thus, if \( \hat{I}_a \) had been enumerated by \( r_k \) \((k=1,2)\) in the dual system survey, a household eligible to report him by \( r_t \) \((t=3-k)\) would be selected for the quality check survey. The quality check survey ascertains whether or not the \( I_a \) \((a=1,\ldots,N)\) that were enumerated by \( r_k \) \((k=1,2)\) in the dual system survey will also be reported by households eligible to report them by \( r_t \). Obviously, it would be impossible to select quality check households either for the individuals that were enumerated by \( r_k \) \((k=1,2)\) if \( s_{at} = 0 \), or for the individuals that were not enumerated by \( r_k \) because \( s_{aik} = 0 \).

The selection of individuals in the quality check survey is represented by the Bernoulli variable

\[ \hat{\epsilon}_{aik} = \begin{cases} 1 & \text{if } \hat{I}_a \text{ is linked to and enumerated at } H_i \text{ by } r_k \text{ (}k=1,2\text{))} \\ 0 & \text{otherwise.} \end{cases} \]

Assuming that (1) the probability of being enumerated by a particular counting rule is the same in the dual system survey and in the quality check survey, and (2) the chances of being enumerated in both surveys are independent, we have

\[ E(\hat{\epsilon}_{aik} | \hat{\epsilon}_{aik}) = P_t \hat{\epsilon}_{aik} \]

and

\[ E(\hat{\epsilon}_{aik}) = \begin{cases} f_k P_k P_t \delta_{aik} & \text{if } s_{at} > 0 \\ 0 & \text{otherwise.} \end{cases} \]

Finally, we have two estimators of the \( N_{t2} \), previously defined by (1) as the number of the \( I_a \) \((a=1,\ldots,N)\) eligible to be enumerated by \( r_1 \) and \( r_2 \).

\[ \hat{N}_{12k} = \frac{M}{m} \frac{M}{k=1} \sum_{i=1}^{M} \alpha_i \sum_{a=1}^{N} \hat{\epsilon}_{aik}/s_{aik} \text{ (}k=1,2\text{)} \]

Under the assumptions specified in the preceding paragraph,

\[ E(\hat{N}_{12k}) = \frac{N_{12} P_1 P_2}{N} \text{ (}k=3-k\text{)} \]

**ESTIMATOR \( \hat{N} \)**

The dual system network estimator of \( N \) is

\[ \hat{N} = \frac{\hat{N}_{121} + \hat{N}_{122}}{N} \text{ (}0 \leq \beta \leq 1\text{)} \]

Assuming large samples, and independence of response in the dual system survey and the quality check survey, we have

\[ E(\hat{N}) = E(N_1) E(N_2) = \frac{N_1 N_2}{E(N_{12})} = \frac{N_{12}}{N_{12}} \]

If \( s_{aik} \geq 1 \) \((a=1,\ldots,N)\) \((k=1,2)\), \( N_1 = N_2 = N_{12} = N \), and it follows that \( E(\hat{N}) = N \). On the other hand,
if \( s_1 > 1 \) \((a=1,\ldots,N)\) but \( s_2 = 0 \) for some of the \( t_a \). \( N_1 = N \) but \( N_2 = N_1 - N \). Nevertheless, under these latter conditions it also follows that \( E(N) = N \). Thus, we have shown that \( N \) is a consistent estimate of \( N \) if \( s_a > 1 \) \((a=1,\ldots,N)\) for either or both counting rules, and if the responses in the dual system survey and the quality check surveys are independent.

CONCLUDING REMARKS

The dual system network estimator has the same form as the traditional dual system estimator. For the traditional estimator, \( N_1 \) and \( N_2 \) are based on the outcomes of two separate data systems, a household survey and registration system respectively, and \( N_12 \) is obtained by matching individuals that are reported by both systems. It has been shown that the household survey used by the traditional dual system estimator may be based on either traditional sampling [1] or network sampling [3].

The dual system traditional estimator and the dual system network estimator are subject to correlation bias if the assumption of response independence is not satisfied. The traditional estimator would have correlation bias if the responses to the two data systems were dependent. The network estimator would have correlation bias if responses to the two counting rules were dependent.

The dual system estimators are also subject to other types of nonsampling errors. For instance, the traditional estimator is subject to matching bias if individuals reported by both data systems are not matched or if individuals not reported by both systems are erroneously matched. Matching bias is likely to occur if the variables used for matching are based on poor quality data. The network estimator is subject to address bias if the addresses of the quality check households can not be located. Address bias occurs when the dual survey households provide incomplete or inaccurate addresses for the quality check households [4].

Overall, however, relatively little is known about the error effects of the traditional dual system estimator and even less is known about the error effects of the dual system network estimator.

ACKNOWLEDGEMENT

The author thanks Gad Nathan for his many helpful comments.

REFERENCES


