Journal of Official Statistics
Vol. 3, No. 4, 1987, pp. 477–483

Letters to the Editor

Letters to the Editor will be confined to discussions of papers which have appeared in the Journal of Official Statistics and of important issues facing the statistical community.

Criteria for the Treatment of Nonresponse in Sample Surveys
A reply to Bethlehem and Kersten (1985)

Introduction

Bethlehem and Kersten (1985) showed how the true confidence level of a confidence interval (based on the respondents only) varies with the ratio of the bias and the standard error of the point estimate. They use these results to formulate a confidence level criterion, what they call their "call-back criterion." The callback criterion signals when call-backs should be made and thus guarantees a minimum level of confidence. The underlying assumption is that the bias is too large if it exceeds the standard error of the estimate, i.e., if:

$$|B(\bar{y}_r)/S(\bar{y}_r)| \geq 1 , \quad (1)$$

in which case an intended 95 % confidence interval has a coverage that is less than 83 %.

Bethlehem and Kersten express this criterion for a dichotomous variable in terms of a relation between the population size, N, the sampling fraction, a, the nonresponse rate, Q, and the contrast, C, as in (5) below. The contrast is the difference between the population means of the response and the nonresponse strata, i.e., $|C = \bar{Y}_r - \bar{Y}_{nr}|$, where r denotes the response stratum and nr denotes the nonresponse stratum. Note that $\bar{y}_r$ is an unbiased estimate of $\bar{Y}_r$, but usually a biased estimate of $\bar{Y}$.

Bethlehem and Kersten present contrast curves for a number of contrast values for the case where $n = 0.05N$ and $\bar{Y}_r$ is a proportion equal to 0.5. From the contrast curves, one can see when the nonresponse bias, for a given nonresponse rate and a given contrast, is too large. They then state that their criterion concerns the confidence level only and does not guarantee accurate estimates. Therefore they suggest that the right-hand side of the diagram not be used (p. 292). The confidence level alone does not guarantee accuracy because a confidence interval can indeed be large.

Another criterion

Obviously it is precision that Bethlehem and Kersten have in mind by suggesting not using the right-hand side of the diagram where increasing nonresponse rates reduce the number of observations and thus the precision of the estimates. To take precision into account, I propose another criterion. I call this criterion the "confidence interval length criterion," because it uses, or rather it consists of, a half of a regular confidence interval. The underlying assumption is that the estimation is not accurate enough if

$$t \cdot S(\bar{y}_r)/E(\bar{y}_r) \geq k , \quad (2)$$

where t denotes a quantile of the standard normal distribution function and k is the maximum accepted relative length of half the confidence interval. For example, if $k = 0.05$ then the length of half the confidence interval should not exceed 5 % of $E(\bar{y}_r) = \bar{Y}_r$.

Assuming a sample drawn by simple random sampling of $n = aN$ units of which $n_r = (1\text{-}Q)n$, the standard error for a dichotomous variable can be expressed as:

$$S(y_r \approx \{(1\text{-}a)\bar{Y}_r(1\text{-}\bar{Y}_r)/(aN(1\text{-}Q))\}^{1/2}. \quad (3)$$

Furthermore, the bias can be expressed as:

$$|B(y_r)| = Q|(\bar{Y}_r - \bar{Y}_{nr})| = QC. \quad (4)$$

By inserting (3) and (4) into (1) and (2), the two criteria can be expressed in the following ways:

$$N \geqslant (1/a\text{-}1)\bar{Y}_r(1\text{-}\bar{Y}_r)/(C^2Q^2(1\text{-}Q)), \quad (5)$$

and

$$N \leqslant (t/k)^2\{(1/a\text{-}1)(1\text{-}\bar{Y}_r)/\bar{Y}_r\}(1\text{-}Q). \quad (6)$$

(If the finite population correction is ignorable, i.e., $1\text{-}a \approx 1$, the criteria could be expressed in terms of the sample size $n = aN$ instead.)

If the population size and the resulting sample size are large enough to satisfy the inequality (5), then the bias is too large compared to the standard error to guarantee an acceptable level of confidence. If the size of the population is small enough to fulfill inequality (6), the relative length of the confidence interval exceeds the required limit k and the corresponding estimate y is not sufficiently accurate. In both cases, it is better not to use the estimate, or, if possible, action should be taken to increase the response rate for that population group. With a given sampling fraction, a given confidence level of the unbiased estimator, and a given value of $\bar{Y}_r$ we can construct the criterion curves for alternative accuracies (k) and assumed contrasts (C). These curves are useful when deciding whether or not an estimate should be used.

Results

In figures 1–4, I assume a constant sampling fraction (a = 5 %) and an intended confidence level corresponding to a 95 % coverage (t = 1.96,) throughout. I then apply Bethlehem and Kersten's method and express the criterion curves in terms of the population size, N, and the nonresponse percentage, $100Q$ for selected values of C and k.

Figure 1 shows the curves for criteria (5) and (6), assuming a $\bar{Y}_r$ equal to 0.5. The dotted curves depict inequality (6) that is an expression of the confidence interval length criterion. If the combination of population size (N) and nonresponse percentage ($100Q$) yields a point *below* the dotted curve, the estimate is inaccurate. The other curves, those that are not dotted, depict the confidence level criteria (5). These are Bethlehem and Kersten's call-back or contrast curves. If the combination of population size and nonresponse percentage yields a point *above* the contrast curve, the true confidence level differs too much from the one stipulated (less than 83 % instead of 95 %).

Figure 1 sheds some light on Bethlehem and Kersten's suggestion of not using the right-hand side of the contrast curve, i.e., poor precision or large confidence interval of the estimate. This suggestion cannot, however, be followed in every case. In the case of $k \geqslant 4\%$ and $C = 1\%$, I believe their proposal to be too stringent. On the other hand, it is more common that $C > 1\%$ and in these cases, the Bethlehem-Kersten suggestion is not strict enough. The acceptance area, i.e., the area below the contrast curve and above the confidence level (dotted) curve is then greatly reduced. For example if $C \geqslant 2\%$ and $k = 2\%$ then a nonresponse rate greater than 25 % can not be accepted.

Now, what happens if $\bar{Y}_r \neq 0.5$? Figures 2 and 3 show the case where $\bar{Y}_r = 0.9$ and $\bar{Y}_r = 0.1$, respectively. Because of symmetry, it is necessary to consider the contrast curve for $\bar{Y}_r < 0.5$ only. For inequality (6), however, we must distinguish between the cases $\bar{Y}_r < 0.5$ and $\bar{Y}_r > 0.5$. Now we can see that as $\bar{Y}_r$ approaches 1, (Fig. 2) the better Bethlehem and Kersten's approach becomes. In this case the confidence level criterion is more important than the interval length criterion.

On the other hand, as $\bar{Y}_r$ approaches zero (Fig. 3), the interval length criterion becomes more important than the confidence level criterion. In this case, Bethlehem and Kersten's proposal is invalid.

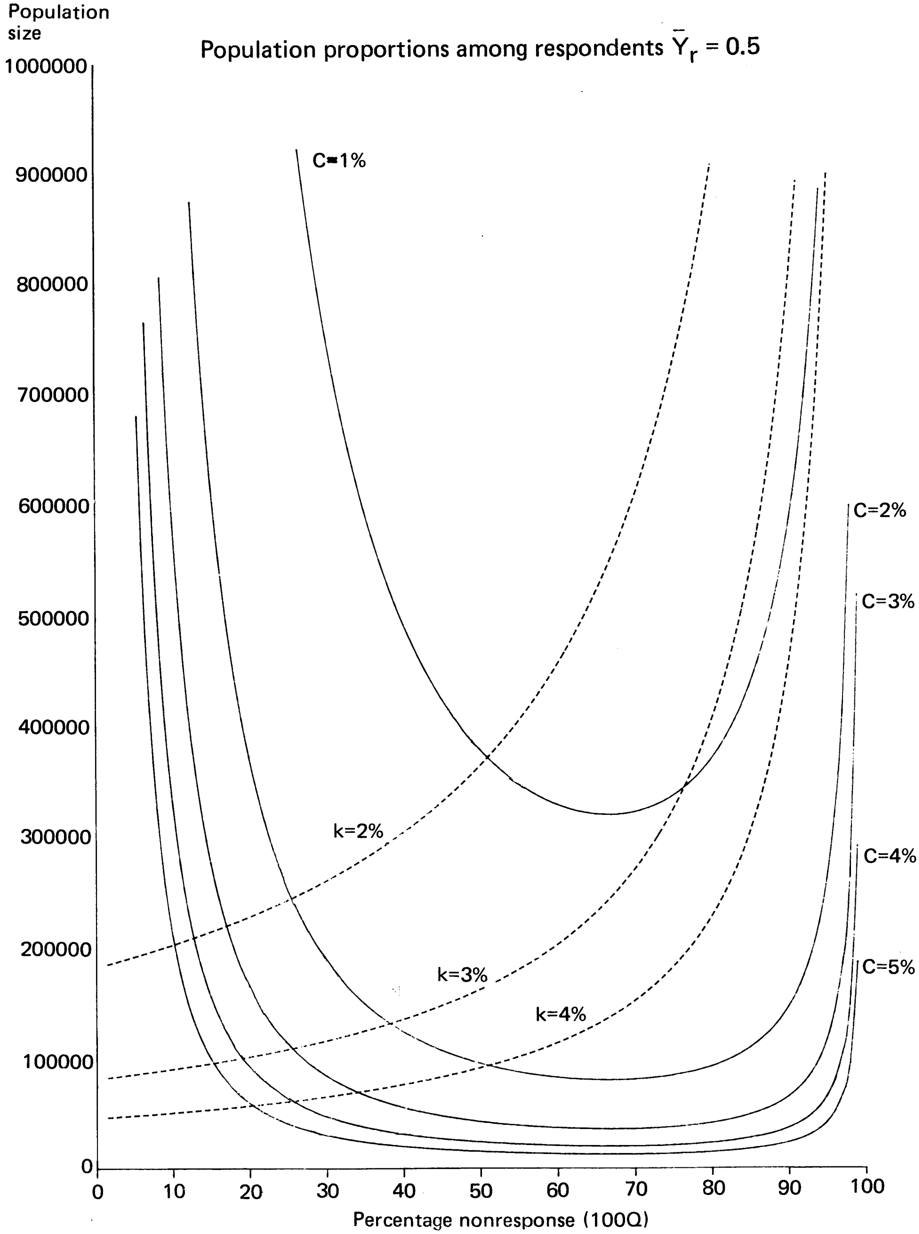

Fig. 1. *Curves for* $\bar{Y}_r = 0.5$ *and selected values of* $C = |\bar{Y}_r - \bar{Y}_{nr}|$ *and* $k = 1.96\, S(\bar{y}_r)/\bar{Y}_r$. *Sample size* $n = 0.05N$

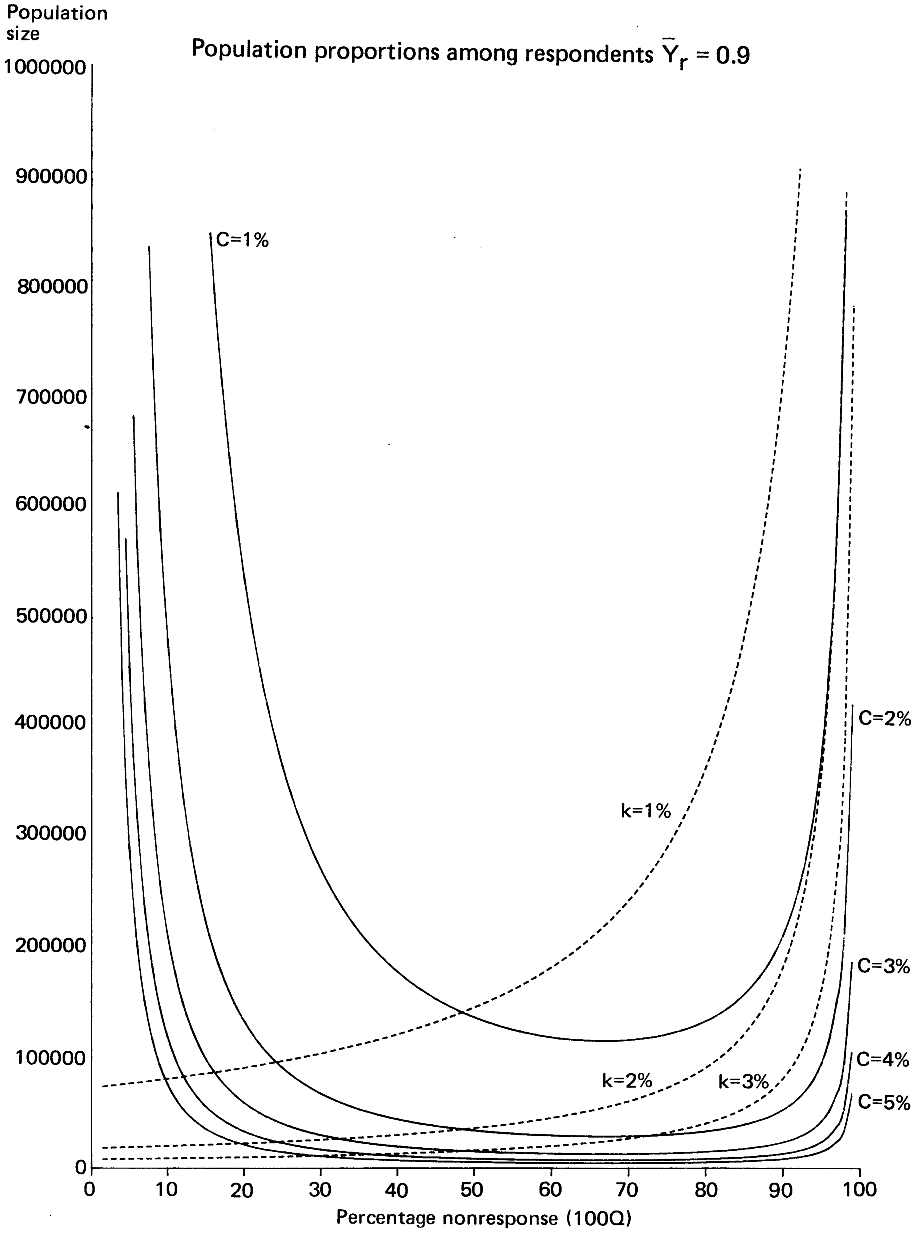

Fig. 2. Curves for $\bar{Y}_r = 0.9$ and selected values of $C = |\bar{Y}_r - \bar{Y}_{nr}|$ and $k = 1.96\, S(\bar{y}_r)/\bar{Y}_r$. Sample size $n = 0.05N$

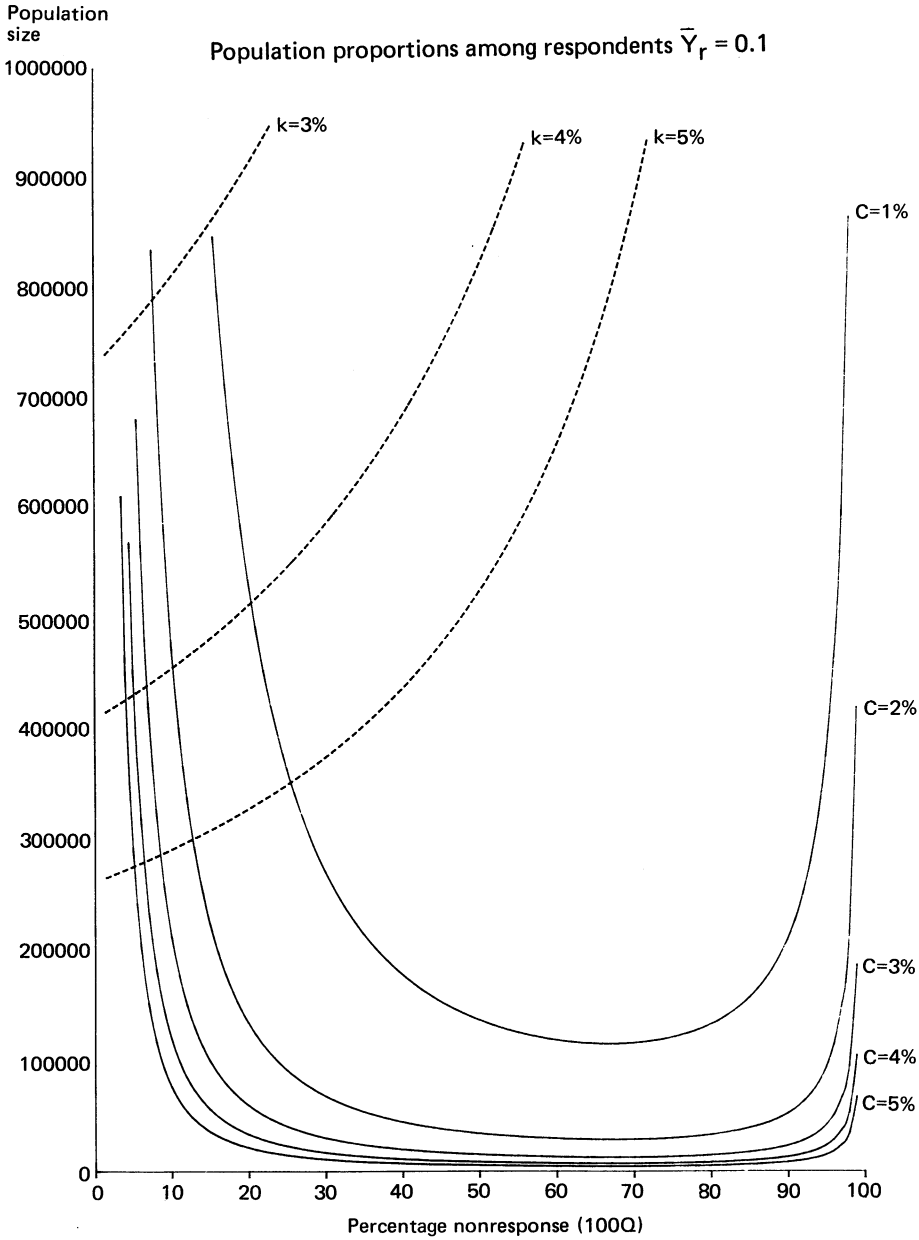

Fig. 3. Curves for $\bar{Y}_r = 0.1$ and selected values of $C = |\bar{Y}_r - \bar{Y}_{nr}|$ and $k = 1.96\, S(\bar{y}_r)/\bar{Y}_r$. Sample size $n = 0.05N$

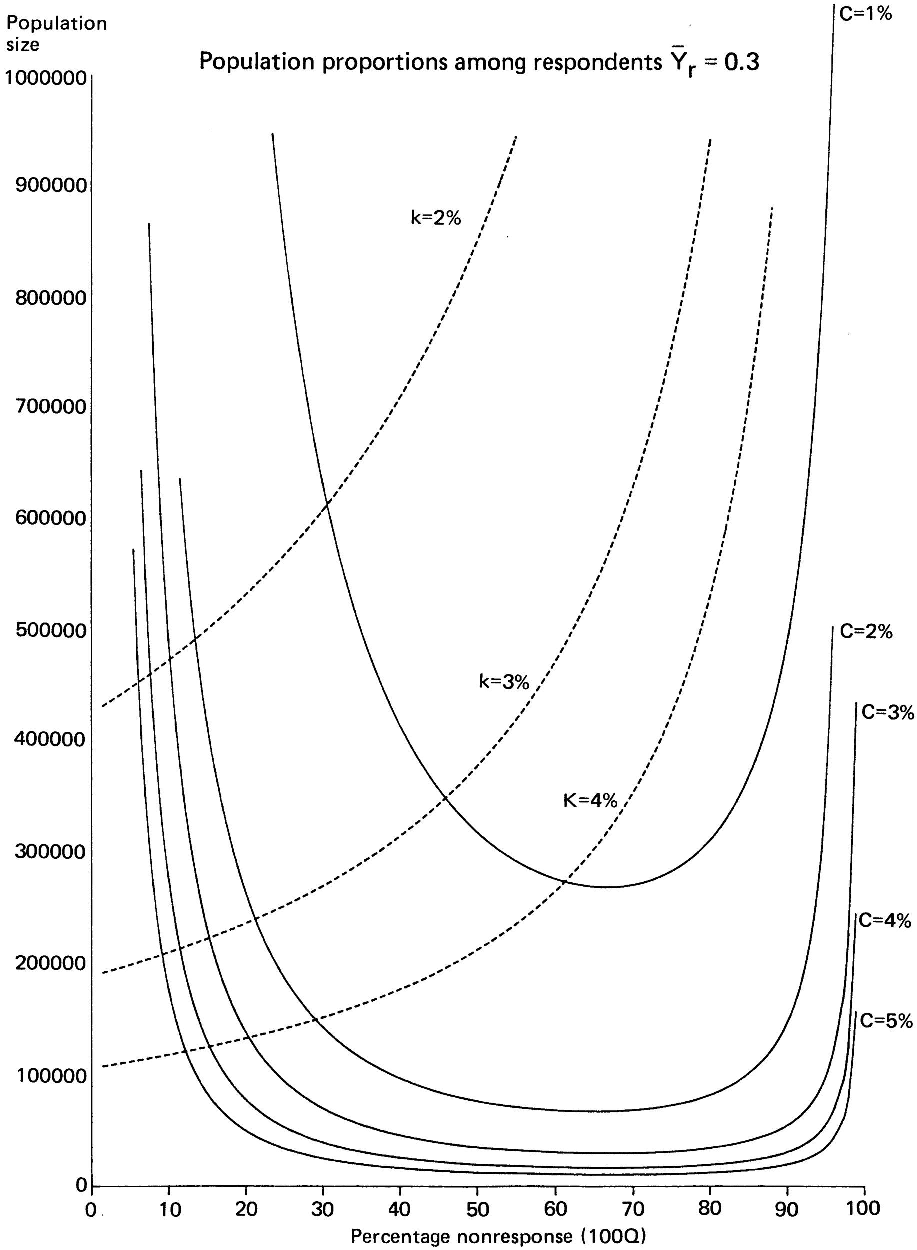

Fig. 4. Curves for $\bar{Y}_r = 0.3$ and selected values of $C = |\bar{Y}_r - \bar{Y}_{nr}|$ and $k = 1.96\,S(\bar{y}_r)/\bar{Y}_r$. Sample size $n = 0.05N$

Conclusions

In the majority of cases, $\bar{Y}_r$ will probably be less than 0.5, and a Y_r that is equal to 0.3 is quite typical. In Figure 4 we see that even if the contrasts are relatively small, e.g., C = 3 %, the number of call-backs increases notably. Thus even if precision is not taken into account, nonresponse rates exceeding 20 % present problems.

If one then takes precision into account, the problems worsen. For example, if k = 2 % and C = 3 %, then nonresponse rates of 10 % or more will not be acceptable.

I believe that this analysis shows that, in general, one cannot expect accurate interval estimates in the presence of nonresponse rates exceeding 30 %. This however is only true, if there is a contrast.

Reference

Bethlehem, J.G. and Kersten, H.M.P. (1985): On the Treatment of Nonresponse in Sample Surveys. Journal of Official Statistics, 1, 3, pp. 287–300.

Roland Gnoss
Johannes-Lüft-Strasse 6
D-6500 Mainz 42
West Germany

Corrigendum

Johnson, E.G. and King, B.F. (1987): Generalized Variance Functions for a Complex Sample Survey. Journal of Official Statistics, Vol. 3, No. 3, pp. 235–250.

Exhibit 6, p.246, was drawn incorrectly. The central rectangle of the histogram should show a frequency of 373 instead of 273, and the height of the rectangle and the vertical scale should be changed accordingly.