ON A PROBABILITY MECHANISM TO ATTAIN AN ECONOMIC BALANCE BETWEEN THE RESULTANT ERROR OF RESPONSE AND THE BIAS OF NONRESPONSE

By

W. EDWARDS DEMING

GRADUATE SCHOOL OF BUSINESS ADMINISTRATION
NEW YORK UNIVERSITY

Reprinted from

THE JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION,
Vol. 48, December 1953: pp. 743-772

Reprinted from the JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
December, 1953, Vol. 48, pp. 743-772

ON A PROBABILITY MECHANISM TO ATTAIN AN ECONOMIC BALANCE BETWEEN THE RESULTANT ERROR OF RESPONSE AND THE BIAS OF NONRESPONSE

W. EDWARDS DEMING
New York University

The author postulates a probability mechanism for the simultaneous production of the bias of nonresponse and for the variance of response. The nonresponse arises from a graded series of classes of the members of the universe to be sampled. The classes range from an impregnable core of no possible response, on up to a class of complete response. Nonresponse arises from two sources, not at home, and refusal. Refusals are of two kinds, permanent and temporary. The variation in the amount of time spent at home, and the variation in the firmness of the temporary refusal, produce the graded series of classes. The bias of nonresponse arises from the variation of any characteristic from one class to another. The variance of response arises from the variation of any characteristics from one member to another within a single class, and from the random variation in the number of responses therefrom.

An increase in the size of the initial sample or a more efficient method of selection will decrease the variance of response, but will have no effect on the bias of nonresponse. Successive recalls, on the other hand, decrease the bias of response, and are more effective than an increase in the size of the sample or a more efficient method of selection in decreasing the root-mean-square error which arises from both nonresponse and from the variation of response.

The results show that without recalls, it is hazardous to put any confidence in the result, no matter how big the sample, even when the variation in the measured characteristic is only two-fold from the class of lowest response to the class of highest response.

With the levels of response assumed here (taken from average urban experience), and with an estimate formed by summing up the initial call and the recalls, the first two recalls effect together about a 50% reduction in the initial bias of nonresponse. Further recalls continue to be productive. In fact, with this method of estimation, each recall added to a sampling plan, even to six recalls, actually increases the amount of information obtained for each dollar expended on interviewing.

Even with three recalls, and with only a two-fold variation from the class of lowest response to the class of highest response, an initial sample bigger than the equivalent of from

A further purpose of the paper is to compare the results and the costs of recalls with the alternative Politz plan.

CRITERION FOR THE OPTIMUM PLAN

We now define the root-mean-square error. The criterion to be adopted here for the optimum plan is that it shall deliver a prescribed mean square error at minimum cost. The root-mean-square error (to be abbreviated r-m-s error hereafter) of any plan of survey will by definition denote the hypotenuse of a right triangle, one leg of which is the bias of the nonresponse that arises from the plan, and the other leg of which is the standard error of the plan (see Fig. 1). Different plans will have different triangles. By definition, the criterion for the optimum plan is that it shall give a shorter hypotenuse than any other plan will give for the same cost; or, alternately, a plan is optimum if it, among all possible plans, will deliver a prescribed length of hypotenuse at the lowest cost. One plan is "better" than another if it will yield a

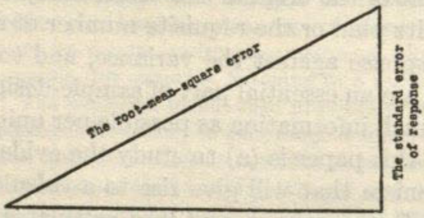

FIGURE 1. Any plan of survey will possess a bias of nonresponse and a standard error of response. The right angle addition of the two forms the root-mean-square error of the particular plan.

shorter hypotenuse than the other, for the same cost. There are a number of nonsampling errors in all surveys, whether complete or sample.[6] The bias of nonresponse is only one of them. It exists, of course, in complete counts as well as in samples. In fact, the conclusions to be reached at the end will point to some drastic re-orientation of the effort expended on complete counts. Both the bias of nonresponse and the error of sampling exist in sample surveys. These are the two errors that within any particular framework of design of sampling, interviewing, and questioning, are direct functions of the size of the sample and of the number of recalls.

[6] A list of such errors with discussion is contained in Chapter 2 of Deming's *Some Theory of Sampling* (John Wiley, 1950); and in an article entitled "On errors in surveys," *American Sociological Review*, ix (1944) 359–69.

As one seldom knows the resultant magnitude of all the nonsampling errors, and as they vary from one survey to another, the most sensible magnitude to aim at for the r-m-s error of the combination of sampling and of nonresponse (the hypotenuse in Fig. 1) will vary likewise. One might aim at a r-m-s error of 7% in one survey, at 10% in another, and at 20% in another. Even with unlimited expenditure to reduce the r-m-s error to very low proportions, other errors will still be present unless funds are diverted to reduce them also.

QUANTIFICATION OF THE PROBLEM

The probability mechanism or model will now be described. The population to be sampled will be divided into six classes, according to the average proportion of interviews that will be completed successfully out of 8 attempts. The classes will be designated by 0, 1, 2, 4, 6, 8 to denote 0, 1, 2, 4, 6, 8 interviews completed, on the average, out of 8 attempts. These figures will often appear as subscripts to various other symbols. Six classes will be sufficient: more classes would not alter the results enough to warrant the extra labor.

We assume that under the conditions specified for any particular survey, failure to obtain an interview may arise from a multitude of causes, which are manifest as not at home and refusal. We assume that people that refuse are of two kinds, those that give permanent refusals and those that give temporary refusals. People that give permanent refusals will never respond to any kind of treatment (they are a part of Class 0 defined more explicitly later). People that give temporary refusals are the kind that will refuse sometimes but will grant interviews at other times or to other interviewers. An example of a temporary refusal is a case where the wrong interviewer called, or the right one called at the wrong time—woman bathing the baby, indisposed, family at dinner, etc. An interview might have been obtained with better luck in timing, or better luck in the selection of the interviewer.

Class 0 contains the stubborn core of permanent impregnable refusals, plus the people who are never at home, gone to Florida, etc., or who are drunk when you do finally find them, or who turn out to be incapacitated otherwise and can not possibly give meaningful answers. At this moment we may note that the magnitude of this class varies widely, dependent on the type of information called for by the survey, and on the procedure of getting it. In a census, when people are away, or refuse, or are incapable of giving information, a good share of the required information can usually be obtained from neighbors, and is, although information on income must usually in such cases be left unanswered. Thus Class 0 in a count of the number of inhabitants only is doubtless well below 1%, being reduced by the cooperation of neighbors. But in surveys whose express purpose is income, expenditures, savings, medical history, the neighbors are unable to help, and Class 0 is bigger. I assume it to be 5% in the calculations to be presented here.

At the other extreme is Class 8, the people who 8 times out of 8 are at home and answer the questions. Moving inward from the soft outer shell (Class 8) toward the impregnable core, we encounter layers of increasing density. In Classes 6, 4, 2, 1 are the temporary refusals plus the people who are not home all the time. In Class 6 an interviewer will be successful at finding the respondent at home and in getting an interview, on the average, 6 times out of 8; in Class 4, 4 times out of 8; etc.

Thus, we have not merely responding units and nonresponding units. Neither have we merely an overall proportion of response nor of nonresponse, but rather, response and (except for Class 8) nonresponse from each of several classes. We have not a mean value of some characteristic for the responses and some other value for the nonresponses; instead, each class possesses a mean and a variance. We are concerned with the cumulative results from all classes.

THE PATIENT MEAN

We define the "patient mean" as

$$a^* = \frac{\sum_1^8 p_i a_i}{\sum_1^8 p_i} = \frac{\sum_1^8 p_i a_i}{q_0} \tag{1}$$

wherein a_i is the mean value per sampling unit of some particular characteristic (rent, number of people employed, or something else) in Class i, and p_i is the proportion in this class. The patient mean will be the datum from which we reckon the biases in later calculations, and the unit in which we shall measure the bias and the root-mean-square error of any plan. It is the result of calling back patiently ad infinitum on all the people in Classes 1, 2, 4, 6, 8. The members of Class 0 will also be included in the recall because in practice we have no way of separating them out; but as they yield no response, they contribute nothing to the patient mean.

THE INITIAL SAMPLE (ATTEMPT I)

The treatment will be simplified by the assumption that the initial sample is the mere drawing of n names from a list of N names (the *frame*). A more complex plan will cause no important modification in the conclusions with respect to the necessity for recalls, nor with respect to the number of recalls required for the most economical plan. It will not modify seriously the comparison with the Politz plan. It will, however, change the absolute figures on cost, but these are not the aim of this study; they are auxiliary only. By further assumption the frame will be so large compared with the sample that the multinomial term

$$P = \frac{n!}{n_0!n_1!n_2! \cdots n_8!} p_0^{n_0}p_1^{n_1}p_2^{n_2} \cdots p_8^{n_8} \tag{2}$$

gives the probability that in the initial sample (Attempt I), there will be n_i names in Class i. n is the size of the initial sample. n_i is a random variable; p_i and n are constants, satisfying the equations

$$\sum_0^8 n_i = n \tag{3}$$

$$\sum_0^8 p_i = 1. \tag{4}$$

If the sample (n) is as great as 10 per cent or more of the frame, the variances and the biases to be computed should be reduced approximately by the factor $1-n/N$, in practice this reduction will be of negligible importance.

When the returns from the initial call come in, we form from them the numerical average for some particular characteristic and denote it by $x(\mathrm{I})$. According to the particular mechanism postulated, the composition of $x(\mathrm{I})$ will be the fraction

$$x(\mathrm{I}) = \frac{\text{Sum of all the numerical values in the responses of Attempt I}}{\text{Number of responses in Attempt I}}. \tag{5}$$

If we were able to separate the returns by class, this would appear as

$$x(\mathrm{I}) = \frac{\sum R_i x_i}{\sum R_i} \quad \begin{array}{l}\text{[Here and hereafter, sums will run over all classes}\\ \text{except 0, unless indicated otherwise]}\end{array} \tag{5a}$$

wherein R_i represents the number of responses from Class i, and x_i

represents the mean of the R_i responses. Both R_i and x_i are random variables. Their expected values are

$$Ex_i = a_i \tag{6}$$

$$ER_i = n\pi_i p_i \tag{7}$$

where

$$\pi_i = \frac{i}{8} \cdot \tag{8}$$

The variance of x_i will be

$$\text{Var } x_i = \frac{\sigma_i}{n\pi_i p_i}\left(1 + \frac{1 - \pi_i p_i}{n\pi_i p_i}\right), \tag{9}$$

wherein σ_i is the standard deviation of the particular characteristic in Class i. In what follows we shall drop terms in $1/n^2$; hence we shall have no further use for the term $(1-\pi_i p_i)/n\pi_i p_i$ in the last equation.

The quantity $x(\text{I})$ in Equation 5 is a random variable. Under the assumed probability its expected value will be

$$E(\text{I}) = \frac{G}{H}, \tag{10}$$

and its variance will be

$$\text{Var (I)} = \frac{8}{nH^2} \sum ip_i\{\sigma_i^2 + [a_i - E(\text{I})]^2\}, \tag{11}$$

where for convenience

$$G = \sum ip_i a_i \tag{12}$$

$$H = \sum ip_i. \tag{13}$$

The derivation of Equations 10 and 11 is simple in the light of certain well-known principles of sampling. Let each sampling unit possess 8 cells , each one NR or R (NR for no response, R for response) according to the following distribution:

Class 0,	8 NR,	0 R
Class 1,	7 NR,	1 R
Class 2,	6 NR,	2 R
Class 4,	4 NR,	4 R
Class 6,	2 NR,	6 R
Class 8,	0 NR,	8 R

Now when we draw a sample, we in effect draw first a sampling unit, which will belong to one of the above classes. Next, we draw 1 of its 8 cells at random to determine whether we get a response. If we draw an R-cell (a response), we write down the number x_{ij}, which will be a random variable, the same for all the R-cells of an individual, but varying from one individual to another. If we draw an NR-cell (no response), we make no record at all. The probability of getting a response in the double drawing (first, an individual sampling unit; second, a cell) is $\pi_i p_i$, which is only the expected proportion of all the responses that will fall in Class i.

The mean of the entire set of responses in the frame will be

$$\mu_R = \frac{\sum \pi_i p_i a_i}{\sum \pi_i p_i} = \frac{G}{H} \tag{14}$$

and their variance will be[7]

$$\sigma^2 = \frac{\sum \pi_i p_i [\sigma_i^2 + (a_i - \mu_R)^2]}{\sum \pi_i p_i}. \tag{15}$$

The double drawing is a random procedure in which each cell has the same probability as any other in the entire frame. The mean of the returns of a sample will therefore give an unbiased estimate of the mean of the entire set of responses; but this is only a restatement of Equation 10. The expected number of responses in a sample of n is $n\sum \pi_i p_i$, wherefore the variance of a sample of n will be very closely equal to $\sigma^2/n\sum \pi_i p_i$; but this is only a restatement of Equation 11. And thus Equations 10 and 11 are established.[8]

The bias in the expected result $E(\mathrm{I})$ of Attempt I will be defined as

$$B(\mathrm{I}) = E(\mathrm{I}) - a^*. \tag{16}$$

The mean square error of $x(\mathrm{I})$ will then be

$$\mathrm{Mse}\ (\mathrm{I}) = \mathrm{Var}\ (\mathrm{I}) + B^2(\mathrm{I}). \tag{17}$$

If Figure 1 were drawn for Attempt I, the two terms on the right of this equation would be the squares of the two legs of the triangle, and the left-hand member would be the square of the hypotenuse.

---

7 This is the formula for the variance of a composite universe; see, for example, the author's *Some Theory of Sampling* (John Wiley & Sons, 1950), pp. 58 and 59.

8 My colleague Dr. Benjamin J. Tepping discovered this simple way of deriving Equations 10 and 11. He furnished also algebraic proofs, but they seem not to be required.

ATTEMPTS II, III, IV

The nonresponses left over from the first attempt form a new frame. The sampling plan may prescribe 0, 1, 2, or more recalls on a sample of these nonresponses.

The 1st recall will be identified here as Attempt II. The 2d and 3d recalls will be Attempts III and IV.

The determination of the optimum fraction (y) of the nonresponses of Attempt I to draw for recall will be a subject for investigation in a later paragraph.

The bias of nonresponse arises from Classes 1, 2, 4, 6. Each successive attempt digs deeper into the lower classes, and diminishes the relative proportions that remain in the upper classes. Class 8 is in fact wiped out in Attempt I. In this way the combination of successive attempts pushes the accumulated result closer and closer to the patient mean a^*.

We assume that each attempt picks up a random sample of the nonresponses in each class. This is not what happens, but it is probably impossible to put down an equation for what actually happens. The interviewers use ingenuity. They find out from neighbors when the people now absent will be at home. They make observations: they make appointments. They hold conferences to decide which one of them might best succeed in breaking down a refusal. Working for and working against the interviewers is some softening and also some hardening of the hearts of people who refused at an earlier call. I have seen them both. The net result is probably that the recalls are less costly (as Houseman says) than I assume in Table 3, and more successful than this theory indicates. If so, then the recommendations for recalls are even stronger than one may conclude from this theory alone.

Equations 10 and 11 apply also for the results of Attempts II, III, IV, if n is treated in any attempt as the number of interviews attempted, and if p_i in Equations 10–13 is replaced by:

$$(1-\pi_i)\, p_i / \sum (1 - \pi_i) p_i \qquad \text{Attempt II}$$

$$(1-\pi_i)^2\, p_i / \sum (1 - \pi_i)^2 p_i \qquad \text{Attempt III}$$

$$(1-\pi_i)^3\, p_i / \sum (1 - \pi_i)^3 p_i. \qquad \text{Attempt IV}$$

Class 8 contributes nothing to these sums, being wiped out by the factor $1-\pi_i$ which is 0 when $i=8$.

EQUATIONS FOR THE COMBINATION OF ATTEMPTS

If the plan of survey calls for two recalls, we combine the results of Attempts I, II, III. With an obvious extension of notation, the result of this combination will be

$$x(\mathrm{I}+\mathrm{II}+\mathrm{III}) = u_{\mathrm{I}}x(\mathrm{I}) + u_{\mathrm{II}}x(\mathrm{II}) + u_{\mathrm{III}}x(\mathrm{III}), \tag{18}$$

where u_{I}, u_{II}, u_{III} are weights. If R_{I}, R_{II}, R_{III} are the responses in the three separate attempts, then

$$u_{\mathrm{I}}, u_{\mathrm{II}}, u_{\mathrm{III}} = \frac{R_{\mathrm{I}}, R_{\mathrm{II}}, R_{\mathrm{III}}}{R_{\mathrm{I}} + R_{\mathrm{II}} + R_{\mathrm{III}}} \cdot \tag{19}$$

For the expected value of $x(\mathrm{I}+\mathrm{II}+\mathrm{III})$ we may write with sufficient approximation

$$E(\mathrm{I}+\mathrm{II}+\mathrm{III}) = w_{\mathrm{I}}E(\mathrm{I}) + w_{\mathrm{II}}E(\mathrm{II}) + w_{\mathrm{III}}E(\mathrm{III}), \tag{20}$$

wherein w_{I}, w_{II}, w_{III}, are the expectations of u_{I}, u_{II}, u_{III}. Formally, with sufficient approximation,

$$w_{\mathrm{I}}, w_{\mathrm{II}}, w_{\mathrm{III}} = \frac{\sum ip_i,\ \sum ip_i(1-\pi_i),\ \sum ip_i(1-\pi_i)^2}{\sum ip_i[1 + (1-\pi_i) + (1-\pi_i)^2]} \cdot \tag{21}$$

Before proceeding, we note that

$$\left.\begin{array}{r} u_{\mathrm{I}} + u_{\mathrm{II}} + u_{\mathrm{III}} = 1 \\ w_{\mathrm{I}} + w_{\mathrm{II}} + w_{\mathrm{III}} = 1 \end{array}\right\} . \tag{22}$$

The bias of $x(\mathrm{I}+\mathrm{II}+\mathrm{III})$ of the combined results of Attempts I, II, III will be defined as

$$B(\mathrm{I}+\mathrm{II}+\mathrm{III}) = E(\mathrm{I}+\mathrm{II}+\mathrm{III}) - a^*. \tag{23}$$

The variance of $x(\mathrm{I}+\mathrm{II}+\mathrm{III})$ may be computed as

$$\begin{aligned} \mathrm{Var}\,(\mathrm{I}+\mathrm{II}+\mathrm{III}) = w_{\mathrm{I}}^2\ \mathrm{Var}\,(\mathrm{I}) + w_{\mathrm{II}}^2\ \mathrm{Var}\,(\mathrm{II}) \\ + w_{\mathrm{III}}^2\ \mathrm{Var}\,(\mathrm{III}). \end{aligned} \tag{24}$$

The notation in the above equations can easily be extended or contracted to more or to fewer attempts. For a plan that uses only one recall, we simply drop the symbol III; also the term $(1-\pi_i)^2$ in Equation 21. For a plan that uses three recalls, we annex a term in IV, and replace $(1-\pi_i)^2$ by $(1-\pi_i)^2+(1-\pi_i)^3$.

THE POLITZ PLAN

The Politz plan includes questions to inquire of each person found at home, and who does not refuse, to ascertain whether he was at home last night at this time, the night before last, etc., to cover the 5 nights preceding the interview, 6 nights in all. Each return is given a weight

w_t, the reciprocal of the number of nights at home over the period of 6 successive nights. The result of applying the Politz plan will be the random variable

$$x(P) = \frac{Sw_tR_tx_t}{Sw_tR_t} \tag{25}$$

wherein S denotes the sum over the 6 Politz classes, and wherein R_t and x_t denote the number of responses and their mean value in the Politz class t. $w_t = 6/(1+t)$, where t is the number of nights at home during the preceding 5 nights. w_t, R_t, and x_t are, all random variables.

In each class except Class 8 it is possible for a person to be at home, during the preceding 5 nights, some number of nights other than his average (π_i). Thus, $E\,w_t$ is not the reciprocal of π_i, but takes the values shown in Equation 29. By applying the formula

$$E\frac{u}{v} = \frac{Eu}{Ev}(1 + C_v^2 - \rho C_uC_v) \tag{26}$$

it is possible to find the expected value of $x(P)$ and to show that the Politz correction for not being at home leads to the bias

$$\begin{aligned} B(P) &= Ex(P) - a^* \qquad\qquad [\text{Definition}] \\ &= a_P - a^* - \left\{\left(1 - \frac{1}{n}\right)\frac{1}{V^2}\sum(\pi_ip_i)^2(B_i - A_i^2)(a_i - a_P)\right. \\ &\qquad \left. + \frac{1}{nV^2}\sum \pi_ip_iB_i(a_i - a_P)\right\}. \end{aligned} \tag{27}$$

The terms in the braces are very small numerically, and we accept with sufficient approximation for our purpose,

$$B(P) = a_P - a^* \tag{28}$$

wherein $\pi_i = i/8$, as heretofore, and

$$\begin{aligned} A_i &= Ew_t = E\frac{6}{1+t} \qquad\qquad [\text{For Class } i] \\ &= \sum_{t=0}^{5}\frac{6}{1+t}\binom{5}{t}(1-\pi_i)^{5-t}\pi_i^t \qquad [\text{Assuming that } t \text{ is a binomial variate}] \\ &= \frac{1}{\pi_i}\sum_{s=1}^{6}\binom{6}{s}(1-\pi_i)^{6-s}\pi_i^s \\ &= \frac{1}{\pi_i}[1 - (1-\pi_i)^6], \end{aligned} \tag{29}$$

$$B_i = Ew_t^2 = E\left(\frac{6}{1+t}\right)^2$$

$$= \sum_{t=0}^{5}\left(\frac{6}{1+t}\right)^2\binom{5}{t}(1-\pi_i)^{5-t}\pi_i^t$$

$$= \frac{6}{\pi_i}\sum_{s=1}^{6}\frac{1}{s}\binom{6}{s}(1-\pi_i)^{6-s}\pi_i^s$$

$$= \frac{6}{\pi_i}\left\{\frac{1}{1}\binom{6}{1}(1-\pi_i)^5\pi_i + \frac{1}{2}\binom{6}{2}(1-\pi_i)^4\pi_i^2\right.$$

$$\left. + \cdots + \frac{1}{6}\pi_i^6\right\}, \tag{30}$$

$$a_P = \frac{ESw_tR_tx_t}{ESw_tR_t}$$

$$= \frac{\sum \pi_i p_i A_i a_i}{\sum \pi_i p_i A_i} = \frac{\sum p_i[1-(1-\pi_i)^6]a_i}{\sum p_i[1-(1-\pi_i)^6]}, \tag{31}$$

$$V = \sum \pi_i p_i A_i. \tag{32}$$

The bias of a plan that uses $k-1$ recalls may be written

$$B(\mathrm{I}-K) = \frac{\sum p_i[1-(1-\pi_i)^k]a_i}{\sum p_i[1-(1-\pi_i)^k]} \tag{33}$$

to the same approximation that appears in Equation 28. With $k=2$, for example, this form gives a numerical verification of the bias $B(\mathrm{I+II+III})$ calculated otherwise by Equations 20 and 23.

The variance of the Politz plan is[9]

$$\mathrm{Var}\,(P) = \frac{1}{nV^2}\sum \pi_i p_i B_i[\sigma_i^2 + (a_i - a_P)^2]$$

$$+ \left(1 - \frac{1}{n}\right)\frac{1}{V^2}\sum (\pi_i p_i)^2(B_i - A_i^2)(a_i - a_P)^2. \tag{34}$$

It is worth noting that if we place $A_i = B_i = 1$, the second term vanishes, and the right-hand member reduces precisely to Equation 15, as it should.

---

[9] My equation for the Politz plan differs from the equations given by Politz and Simmons.

SEARCH FOR THE OPTIMUM PLAN

The accumulated mean square error of Attempts I, II, and III will be

$$M(\mathrm{I} + \mathrm{II} + \mathrm{III}) = \mathrm{Var}\,(\mathrm{I} + \mathrm{II} + \mathrm{III}) + B^2(\mathrm{I} + \mathrm{II} + \mathrm{III}). \quad (35)$$

We drop the symbol III for a plan that calls for two attempts, and we annex IV for a plan that calls for four attempts.

Any two plans may be expected to incur different costs and to yield different mean square errors. As agreed at the beginning, a plan is optimum if its cost is less than that of any other plan that will yield the same mean square error. This is a matter of numerical calculation.

Numerical assignments to the various fundamental magnitudes (p_i, a_i, σ_i) will occur two sections ahead.

We have one other task—the determination of the optimum fraction y, a subject for the next section.

DETERMINATION OF THE OPTIMUM FRACTION OF NONRESPONSES TO INCLUDE IN THE RECALLS

Let y denote the fraction sought. We remind the reader that Attempt III will be a canvass of all the nonresponses that remain from Attempt II, and that Attempt IV will be a canvass of all the nonresponses that remain from Attempt III. There is thus only the one fraction y to determine.

The mean square error (M) of the accumulated result of any number of attempts may be written in terms of n and y as

$$M = A + B/n + C/ny, \quad (36)$$

the cost of which is

$$Y = Dn + Eny, \quad (37)$$

A, B, C, D, E are constants. As before, n is the initial sample for Attempt I. By differentiation it can be shown that, for a fixed value of Y, the minimum in M occurs when

$$y^2 = \frac{CD}{BE} \cdot \quad (38)$$

This result is independent of n, hence it holds for any initial size of sample.

The equation for y^2 just given contains D and E only in the ratio $D{:}E$, which shows that y does not depend directly on the absolute magnitudes to be assumed for the costs in Attempt I and later, but rather on the ratio of these costs. And as y will be proportional to $\sqrt{D:E}$, y is relatively insensitive to the ratio assumed for $D:E$. Moreover, y is not dependent on the absolute magnitudes of the a_i, but on their ratio to any one of them, or to a^*, because B and C occur only in the ratio $B:C$.

Table 4 shows the optimum values of y obtained from Equation 38; also the values selected for actual use in the calculations. The fraction y obviously varies slowly with the number of recalls. To simplify the required calculations I have set $y=3/5$ for all plans with the first set of a_i; and $y=1/4$ for all plans with the second set of a_i.

It may be of interest to note that the removal of the bias of nonresponse by recalls is independent of the fraction y. It is not necessary to recall on the optimum fraction, nor on any other particular fraction, so far as the bias of nonresponse is concerned. However, as y decreases, the cost goes down but the variance and the r-m-s error increase, so it is wise not to make y too small. The optimum fraction, if it can be predicted on experience, or some approximation thereto, will guide one close to the minimum r-m-s error for any permissible cost of interviewing.

NUMERICAL MAGNITUDES ASSUMED

In order to make numerical calculations and to derive conclusions therefrom with respect to the most economical design of surveys, it is necessary to assume some numerical magnitudes for the p_i, σ_i; also for the costs. Unfortunately, no set of numerical magnitudes can be typical of all conditions met in the field. I may interject the reminder that every question on a questionnaire has not only its own particular values of a_i and of σ_i, but of p_i as well, even within the same survey, because some questions receive better cooperation than others. The best that one can do is to make numerical assumptions that fit some of the conditions met in practice, and to infer from the equations the range of validity of the conclusions.

The basic numerical assumptions are in Table 1. The expected number of interviews, of responses, and of nonresponses, are shown in Table 2. The response rates (the p_i) assumed here are intended to assimilate average urban experience on a question of moderate difficulty; and without making them responsible for the final choice, I wish to thank Messrs. Lester R. Frankel and Robert Weller of the Alfred Politz Research organization for their help and interest in choosing these particular values.

Fortunately, there is a great deal more generality in the two sets of a_i than may be apparent at first sight, for one may transform either one

TABLE 1

NUMERICAL VALUES ASSUMED

Property and symbol		Class 0	1	2	4	6	8	1–8
Proportion p_i		.05	.10	.10	.20	.25	.30	.95
Mean value of the measured characteristic	a_i (1st)	xxx	2.00	1.75	1.50	1.25	1.00	$a^*=1.355\ 263$
	a_i (2d)	xxx	.10	.20	.40	.60	1.00	$a^*=0.589\ 474$
Standard deviation σ_i		xxx	Same as a_i in both sets of a_i					

TABLE 2

THE EXPECTED SIZES OF SAMPLE IN THE VARIOUS ATTEMPTS, BASED ON AN INITIAL SAMPLE OF n IN ATTEMPT I. HERE THE SUMS RUN OVER ALL CLASSES, 0 TO 8

Attempt	Interviews	Responses	Nonresponses
I	n	$n\sum\pi_i p_i$	$n\sum(1-\pi_i)p_i$
II	$n_{\text{II}}=ny\sum(1-\pi_i)p_i$	$ny\sum(1-\pi_i)\pi_i p_i$	$ny\sum(1-\pi_i)^2 p_i$
III	$n_{\text{III}}=ny\sum(1-\pi_i)^2 p_i$	$ny\sum(1-\pi_i)^2\pi_i p_i$	$ny\sum(1-\pi_i)^3 p_i$
IV	$n_{\text{IV}}=ny\sum(1-\pi_i)^3 p_i$	$ny\sum(1-\pi_i)^3\pi_i p_i$	$ny\sum(1-\pi_i)^4 p_i$
V	$n_{\text{V}}=ny\sum(1-\pi_i)^4 p_i$	$ny\sum(1-\pi_i)^4\pi_i p_i$	$ny\sum(1-\pi_i)^5 p_i$
VI	$n_{\text{VI}}=ny\sum(1-\pi_i)^5 p_i$	$ny\sum(1-\pi_i)^5\pi_i p_i$	$ny\sum(1-\pi_i)^6 p_i$
VII	$n_{\text{VII}}=ny\sum(1-\pi_i)^6 p_i$	$ny\sum(1-\pi_i)^6\pi_i p_i$	$ny\sum(1-\pi_i)^7 p_i$
	Numerical values based on an initial sample of $n=1000$		
I	$n=1000$	625.0	375.0
II	$n_{\text{II}}=375.0y$	$126.6y$	$248.4y$
III	$n_{\text{III}}=248.4y$	$60.3y$	$188.1y$
IV	$n_{\text{IV}}=188.1y$	$34.4y$	$153.7y$
V	$n_{\text{V}}=153.7y$	$22.2y$	$131.5y$
VI	$n_{\text{VI}}=131.5y$	$15.6y$	$115.9y$
VII	$n_{\text{VII}}=115.9y$	$11.7y$	$104.2y$

of these sets into almost any other that he may encounter. For example, to discuss a yes-and-no survey in which the proportions of yes vary from 60% in Class 1 to 40% in Class 8, one has only to derive a new value a_i' from an old a_i by setting

$$a_i' = 20 + 20a_i \tag{39}$$

where a_i on the right belongs to the 1st set of a_i in Table 1. Both a_i and $a_i'-20$ have a 2-fold variation from Class 1 to Class 8. The new patient mean is

$$a'^* = 20 + 20a^* \tag{40}$$

where $a^*=1.355\ 263$, the patient mean of the 1st set of a_i, as given in Table 1. The relative bias computed for $a_i'-20$, for any number of attempts, will be precisely the same as the relative bias computed for a_i (Table 5). It follows that the new expected value for any number of attempts will be

$$\begin{aligned} E' &= 20a^* \text{ Rel } B + a'^* \\ &= 47.105 + 27.105 \text{ Rel } B \end{aligned} \tag{41}$$

where Rel B is the relative bias shown in Table 5 for the corresponding number of attempts. An example will occur later (Table 9).

The 2d set of a_i could serve the same purpose by a suitable transformation, but we shall not carry it through.

Thus, in spite of the limitations of any particular set of numerical assumptions, the conclusions to be drawn will warrant some sweeping generalizations.

COSTS

For the costs of making calls (interviewing only) we assume for calculation the following figures:

For Attempt I,	\$3 per call
For later attempts,	\$5 per call
For the Politz plan,	\$4 per name

This amount will cover the cost of weighting and of calling back on the temporary refusals.

Table 3 shows the costs of interviewing derived from the values assumed for the p_i in Table 1, and with the cost per call as mentioned earlier. n is the size of the initial sample, and y is the fraction of the nonresponses left over from Attempt I that constitute the sample for Attempt II.

TABLE 3

COSTS OF INTERVIEWING

Plan	No. of recalls	Cost (dollars)
Attempt I	0	$Y = 3n$
Attempts I+II	$.3750ny$	$3n + 1.8750ny$
Attempts I−III	$.6234ny$	$3n + 3.1172ny$
Attempts I−IV	$.8115ny$	$3n + 4.0576ny$
Attempts I−V	$.9653ny$	$3n + 4.8263ny$
Attempts I—VI	$1.0968ny$	$3n + 5.4839ny$
Attempts I−VII	$1.2126ny$	$3n + 6.0632ny$
Politz	(equivalent to 5 recalls)	$4n$

The actual numerical magnitudes of these costs are not so important as their relative magnitudes. If all the costs were doubled, the cost computed for any plan will be doubled, but the relative costs and the relative merits of the various plans would remain unchanged.

TABLE 4

RESULTS FOR THE OPTIMUM y, AND THE VALUES SELECTED FOR THE CALCULATIONS THAT LED TO TABLES 5 AND 6, AND TO FIGS. 2 AND 3

Plan	1st set of a_i		2d set of a_i	
	y calculated from Equation 38	y selected for calculation	y calculated from Equation 38	y selected for calculation
I−II	.69	3:5	.33	1:4
I−III	.67	3:5	.30	1:4
I−IV	.65	3:5	.28	1:4
I−V	.63	3:5	.26	1:4
I−VI	.61	3:5	.24	1:4
I−VII	.60	3:5	.23	1:4

It should be noted that these costs are for the interviewing only. Considerations of overhead costs, training, and office-work for the different plans must be taken into account before one decides definitely whether one plan is more economical than another.

CONCLUSIONS FROM THE CALCULATIONS

The numerical results of the calculations are in Tables 5, 6, 7, 8 and in Figs. 2 and 3. The biases and r-m-s errors are expressed in units of

TABLE 5

NUMERICAL VALUES OF THE BIASES AND R-M-S ERRORS FOR VARIOUS SIZES OF INITIAL SAMPLE (n); 1st SET OF a_i, $y=.6$. COSTS AT $n=1000$

Plan	I	I+II	I−III	I−IV	I−V	I−VI	I−VII
Rel bias	−.110874	−.075752	−.057302	−.045330	−.036800	−.030383	−.025380
$n=100$							
Rel m-s-e	.025974	.019918	.017628	.016530	.015938	.015594	.015389
Rel r-m-s-e	.161164	.141131	.132773	.128569	.126246	.124876	.124052
$n=200$							
Rel m-s-e	.019133	.012828	.010456	.009293	.008646	.008259	.008016
Rel r-m-s-e	.138322	.113261	.102255	.096402	.092984	.090879	.089532
$n=300$							
Rel m-s-e	.016853	.010464	.008065	.006880	.006215	.005813	.005559
Rel r-m-s-e	.129822	.102294	.089805	.082946	.078834	.076243	.074559
$n=500$							
Rel m-s-e	.015029	.008574	.006153	.004950	.004271	.003857	.003593
Rel r-m-s-e	.122593	.092596	.078441	.070356	.065353	.062105	.059942
$n=1000$							
Rel m-s-e	.013661	.007156	.004718	.003503	.002812	.002390	.002118
Rel r-m-s-e	.116880	.084593	.068688	.059186	.053028	.048888	.046022
$n=2000$							
Rel m-s-e	.012977	.006447	.004001	.002779	.002083	.001657	.001381
Rel r-m-s-e	.113920	.080293	.063253	.052716	.045640	.040706	.037162
$n=3000$							
Rel m-s-e	.012749	.006211	.003762	.002537	.001840	.001412	.001135
Rel r-m-s-e	.112911	.078810	.061335	.050369	.042895	.037577	.033690
$n=5000$							
Rel m-s-e	.012567	.006022	.003571	.002345	.001646	.001216	.000939
Rel r-m-s-e	.112103	.077602	.059758	.048425	.040571	.034871	.030643
Costs at $n=1000$	\$3000	4125	4870	5435	5896	6290	6638

a^*. The base for the bias is the 0-point of the scale for the a_i. The estimation is assumed to be a summation of the initial call and the recalls. The aim is assumed to be the estimation of an average or of a total.

A. Conclusions from the 1st set of a_i, a 2-fold variation from a_1 to a_8: Table 5 and Fig. 2. Conclusions 1, 2, 3, 4, and 5b are independent of the type and size of sample.

1. With no recalls at all (Attempt I only), the minimum relative

r-m-s error attainable is 11%. No sample however big, not even a complete count, can penetrate below this minimum, without recalls.

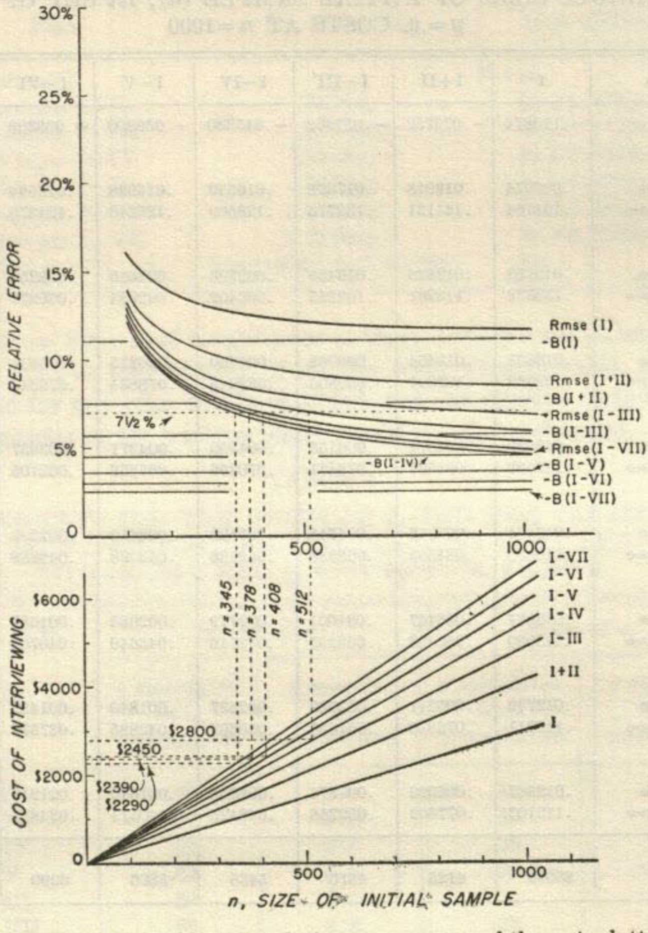

FIGURE 2. The relative bias, the relative r-m-s error, and the cost, plotted against the initial sample-size (n) for various plans, for the 1st set of a_i, in which $a_1 = 2\ a_8$. The curves show the futility of attempting to achieve accuracy by sheer size of sample. Recalls are much more effective. The dashed lines show the size of sample required, and the cost, to yield a relative r-m-s error of $7\frac{1}{2}\%$. The relative biases and the relative r-m-s errors are in units of a^*.

2. With one recall (Attempts I+II), the minimum r-m-s error drops to 7.6%. No sample however big can penetrate below this minimum with only one recall.

3. With 2 recalls (Attempts I+II+III), the minimum r-m-s error drops to 5.7%. No sample however big can penetrate below this minimum with only two recalls.

4. With 3 recalls (Attempts I–IV), the minimum r-m-s error drops to 4.5%. With 4, 5, and 6 recalls, the minimum r-m-s error drops to 3.7, 3.0, and 2.5%.

5. To attain a prescribed r-m-s error of (e.g.) 7½%:

(a) We may use 3, 4, 5, or 6 recalls with initial samples as shown in the accompanying table.

FROM FIG. 2

No. of recalls	Initial sample	Cost
6	345	$2,290
5	378	2,390
4	408	2,450
3	512	2,800

(b) With 0, 1, or 2 recalls we can not attain the prescribed r-m-s error (7½%) with any sample however big.

B. Conclusions from the 2d set of a_i, a 10-fold variation from a_1 to a_8: Table 6 and Fig. 3. Conclusions 6, 7, 8, 9, and 10b are independent of the type and size of sample.

6. With no recalls at all (Attempt I only), the minimum r-m-s error attainable is 24.5%. No sample however big, not even a complete count, can penetrate below this minimum without recalls.

7. With one recall (Attempt I+II), the minimum r-m-s error drops to 15.5%. No sample however big can penetrate below this minimum with only one recall.

8. With 2 recalls (Attempt I+II+III), the minimum r-m-s error drops to 11.3%. No sample however big can penetrate below this minimum with only two recalls.

9. With 3 recalls (Attempts I–IV), the minimum r-m-s error drops to 8.7%. With 4, 5, and 6 recalls, the minimum r-m-s error drops to 6.9, 5.6, and 4.7%.

10. To attain a prescribed r-m-s error of (e.g.) 10%:

(a) We may use 3, 4, 5, or 6 recalls with initial samples as shown in the accompanying table.

From Fig. 3

No. of recalls	Initial sample	Cost
6	210	$ 875
5	245	1,010
4	325	1,380
3	730	2,910

(b) With 0, 1, or 2 recalls we can not achieve the prescribed r-m-s error (10%) with any sample however big.

TABLE 6

NUMERICAL VALUES OF THE BIASES AND R-M-S ERRORS FOR VARIOUS SIZES OF INITIAL SAMPLE (n); 2d SET OF a_i, $y=.25$. COSTS AT $n=1000$

Plan	I	I+II	I−III	I−IV	I−V	I−VI	I−VII
Rel bias	.245190	.155062	.112665	.086955	.069408	.056593	.046815
$n=100$							
Rel m-s-e	.091985	.046455	.032012	.025385	.021763	.019563	.018135
Rel r-m-s-e	.303290	.215534	.178919	.159327	.147523	.139868	.134666
$n=200$							
Rel m-s-e	.076052	.035250	.022353	.016473	.013290	.011383	.010164
Rel r-m-s-e	.275775	.187750	.149509	.128347	.115282	.106691	.100817
$n=300$							
Rel m-s-e	.070741	.031515	.019133	.013502	.010466	.008656	.007507
Rel r-m-s-e	.265972	.177525	.138322	.116198	.102303	.093038	.086643
$n=500$							
Rel m-s-e	.066492	.028527	.016558	.011125	.008207	.006475	.005381
Rel r-m-s-e	.257860	.168899	.128678	.105475	.090592	.080467	.073355
$n=1000$							
Rel m-s-e	.063306	.026286	.014626	.009343	.006512	.004839	.003787
Rel r-m-s-e	.251607	.162130	.120938	.096659	.080697	.069563	.061539
$n=2000$							
Rel m-s-e	.061712	.025166	.013660	.008451	.005665	.004021	.002990
Rel r-m-s-e	.248419	.158638	.116876	.091929	.075266	.063411	.054681
$n=3000$							
Rel m-s-e	.061181	.024792	.013338	.008154	.005383	.003748	.002724
Rel r-m-s-e	.247348	.157455	.115490	.090299	.073369	.061221	.052192
$n=5000$							
Rel m-s-e	.060756	.024493	.013080	.007917	.005157	.003530	.002512
Rel r-m-s-e	.246487	.156502	.114368	.088978	.071812	.059414	.050120
Costs at $n=1000$	$3000	3469	3779	4014	4207	4371	4516

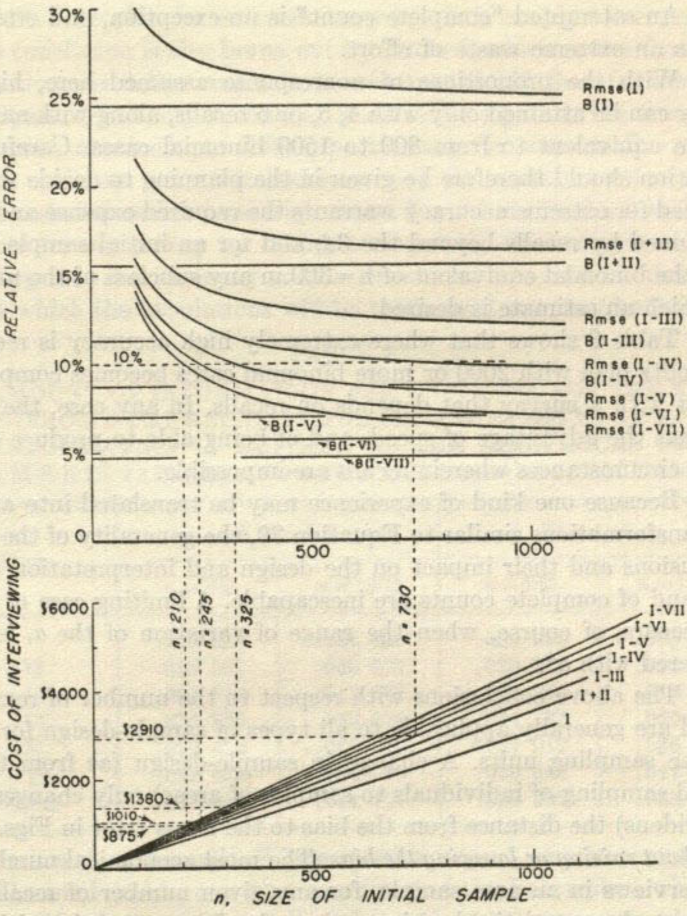

FIGURE 3. The relative bias, the relative r-m-s error, and the cost, plotted against the initial sample-size (n) for various plans, for the 2d set of a_i, in which $a_1 = .1\ a_8$. The curves show the futility of attempting to achieve accuracy by sheer size of sample. Recalls are much more effective. The dashed lines show the size of sample required, and the cost, to yield a relative r-m-s error of 10%. The relative biases and the relative r-m-s errors are in units of a^*.

C. General conclusions

11. Even with three recalls, with the level of response assumed in the calculations (taken from average urban experience), a sample bigger than the binomial equivalent of from 300 to 500 for an estimate of any one class is ineffective and uneconomical. A plan that would reap any real benefit from bigger samples must support 4 or 5 or more recalls.

12. An attempted "complete count" is no exception, and often represents an extreme waste of effort.

13. With the proportions of nonresponse assumed here, high accuracy can be attained only with 4, 5, or 6 recalls, along with an initial sample equivalent to from 800 to 1500 binomial cases. Careful consideration should therefore be given in the planning to decide whether the need for extreme accuracy warrants the required expense and delay occasioned by recalls beyond the 3*d*, and for an initial sample bigger than the binomial equivalent of $n=300$ in any subclass of the universe for which an estimate is desired.

14. Table 8 shows that where extremely high accuracy is required, the Politz plan with 2000 or more binomial cases becomes competitive in cost with a survey that depends on recalls. In any case, the Politz plan has the advantage of speed, and of being able to produce results under circumstances wherein recalls are impossible.

15. Because one kind of experience may be translated into another by transformations similar to Equation 39, the generality of the above conclusions and their impact on the design and interpretation of surveys and of complete counts are inescapable. A limiting case of exception occurs, of course, when the range of variation of the a_i is small compared with a^*.

16. The above conclusions with respect to the number of recalls required are generally applicable to all types of sample-design for drawing the sampling units. A change in sample-design (as from the binomial sampling of individuals to samples of areas) only changes (usually widens) the distance from the bias to the r-m-s error in Figs. 2 and 3, *without raising or lowering the bias*. The most economical number (n) of interviews in an area sample, for any given number of recalls, will for most characteristics be bigger than the figures mentioned in conclusions 11, 13, and 14. The increase may range from 0 on up to sometimes double, depending on the characteristic and the clustering effect of the interviewers' workloads.

IMPACT ON DESIGN

The most impressive feature of the results is the heavy bias of nonresponse, when no provision is made to reduce it, even though there be but a 2-fold variation from a_1 to a_8.

The second most impressive feature is the fact that if nonresponse reaches anywhere near the proportions (p_i) assumed, then when the 0 of the scale of the a_i is not large, we can not afford, except for special justification, to plan for extreme accuracy: it is simply too expensive.

This conclusion is also borne out by Table 7, which shows that more information per dollar comes from a sample of 500 than for a sample of 1000; and that every successive recall shows a gain in the amount of information obtained per dollar, particularly for the smaller sample size. An optimum is not reached even with six recalls. In other words, as we concluded earlier from Figs. 2 and 3 and from Tables 5 and 6, we get more for our money by taking a moderate initial sample and digging deep into it with many recalls. However, many recalls delay the day on which the tabulations will be ready, and one may be forced to

TABLE 7

THE AMOUNT OF INFORMATION PER UNIT COST FOR THE SEVEN PLANS (FROM 0 TO 6 RECALLS), FOR INITIAL SAMPLES OF 500 AND 1000. INFORMATION IS DEFINED AS THE RECIPROCAL OF THE REL M-S-E IN TABLES 5 AND 6. THE COST COMES FROM TABLE 3

Plan	1st set of a_i		2d set of a_i	
	$n=500$	$n=1000$	$n=500$	$n=1000$
I	.044 358	.024 400	.005 013	.005 265
I+II	.056 562	.033 877	.020 216	.010 966
I−III	.066 744	.043 522	.031 954	.018 092
I−IV	.074 326	.052 524	.044 787	.026 664
I−V	.079 422	.060 315	.057 912	.036 501
I−VI	.082 438	.066 519	.070 649	.047 278
I−VII	.083 856	.071 127	.082 302	.058 472

call a halt at 3 or 4 recalls. Where speed is urgent, or where recalls are otherwise inadvisable, one may bear in mind the Politz plan, which offers a rapid solution with recalls only on the temporary refusals.

With the usual method of estimation (pooling the initial call and the recalls) the best way to attain accuracy is to build up the initial response (i.e., to increase p_8). One or two recalls would then be much more effective than they are under the conditions assumed; and bigger samples would also be more effective. Observations on the proper time of day to find certain kinds of people at home in a particular area, and willing to answer questions, plus a skillful introduction and approach so as to cut down refusals, are known to be helpful in this direction.

An attempted complete count is no exception to the conclusions reached. Without a highly successful initial response, followed by some effective number of recalls, 95% of the energy put into a complete

TABLE 8

THE RELATIVE BIAS AND THE RELATIVE R-M-S-ERROR OF THE POLITZ PLAN

1st set of a_i

n	200	300	500	1000	2000	3000	4000	5000
Rel B(P)	−.030 384	−.030 384	−.030 384	−.030 384	−.030 384	−.030 384	−.030 384	−.030 384
Rel Var (P)	.015 566	.010 569	.006 572	.003 575	.002 075	.001 575	.001 376	.001 176
Rel r-m-s-e (P)	.128 410	.107 201	.086 573	.067 067	.054 743	.049 980	.047 424	.045 815

2d set of a_i

n	200	300	500	1000	2000	3000	4000	5000
Rel B(P)	.003 202	.003 202	.003 202	.003 202	.003 202	.003 202	.003 202	.003 202
Rel Var (P)	.016 384	.011 948	.008 401	.005 740	.004 410	.003 965	.003 744	.003 611
Rel r-m-s-e (P)	.139 950	.123 085	.107 717	.094 562	.087 247	.084 658	.083 343	.082 541

count, taken to obtain an estimate for a large area, may be wasted. Size does not atone for nonresponse: this is all too evident from the calculations (Tables 5 and 6; Figs. 2 and 3).

The mechanism adopted here is a device by which experience can be accumulated and pointed toward the attainment of (a) greater accuracy per unit cost, and (b) less waste, through conservation of unproductive effort expended on samples that are too big. Good guesses for the constants p_i, a_i, σ_i can almost always be made on the basis of past experience; and the calculations made with them will indicate a plan not far from optimum. Continued experience will provide improved numerical values for the constants, and continually improved design and interpretation of the results. Without a probability design of some sort, it is difficult to capitalize on experience.

Although the discourse here has been entirely in terms of interviews, the results are equally applicable to surveys in which the initial attempt is made by mail, or in which all attempts are made by mail. Appropriate changes must of course be made in the numerical values of the constants. Thus, if the mail were used for Attempt I, and if interviews were used for the recalls, then the cost D in Equation 38 would be much less than it is when interviews are used in Attempt I, and y will then be smaller. For example, if the cost of a mailed questionnaire were \$.75, and if the cost of an interview on a nonresponse were \$5, then y would reduce to perhaps as low a figure as 1 in 6, depending of course on the other constants in the equation.

One may well wonder what the biases are in surveys that depend only on a mailed survey with a 15% total response, or even 30% or 50%, without calls on the nonresponses. The mechanism adopted here shows that it is a mystery how such results can be worth anything at all.

IMPACT ON METHODS OF ESTIMATION

After the returns from the survey are in, there remains the problem of estimating the mean per sampling unit, and the standard error of this estimate. As the survey does not touch Class 0, it can by itself only produce estimates for Classes 1–6.

The usual practice of combining the various attempts (after weighting Attempt II and higher attempts by the factor $1/y$) may be both misleading and inefficient. A glance at Table 5 or at Figure 2 shows that 41% of the bias still remains after the 3d recall, and that 27% still remains after the 5th. Table 6 and Figure 3 are equally discouraging. The decreasingly slow ascent toward the vertex of 0 bias may explain how easy it is to conclude, incorrectly, that after 3 recalls there is little more bias to squeeze out, and that additional recalls are not worth their cost.

To illustrate the usual procedure, let us make some calculations on a yes and no survey.[10] The proportions of yes in the various classes will range, let us suppose, from 60% in Class 1 down to 40% in Class 8, following the relative values of $20(a_i+1)$ derived from the 1st set of a_i in Table 1. Table 9, calculated with the aid of Equation 41, shows the expected results of combining 2 attempts, 3 attempts, etc. The result that we really need is the patient mean, shown at the bottom of the table as the expected result of continuing the recalls indefinitely. The slow progress of the combined result is obvious; also the need of something better.

TABLE 9

THE EXPECTED PROPORTIONS OF YES, FOR SEVERAL PLANS, COMPUTED BY EQUATION 41. THE PROPORTIONS OF YES RANGE FROM 60% IN CLASS 1 TO 40% IN CLASS 8

Plan	Expected proportion		Bias remaining
	Yes	No	
Attempt I	44.10	55.90	100.0%
I+II	45.05	54.95	68.4
I+II+III	45.55	54.45	51.8
I−IV	45.88	54.12	40.9
I−V	46.11	53.89	33.2
I−VI	46.28	53.72	27.6
I−VII	46.42	53.58	22.8
Infinity	$a'^* = 47.11$	52.89	0

What we need is a way to extract more information from the recalls. A more efficient estimate may be contained in a scheme for extrapolating the results of the various attempts, as proposed by Hendricks.[11] The mechanism proposed here will provide a rational scale for the extrapolation. It may be that the scale proposed by Hendricks is ap-

[10] I am indebted to Dr. Leo P. Crespi and to Mr. Fred W. Trembour of the Reactions Analysis Staff in the Office of the High Commissioner for Germany, who in several conversations with the author brought up questions and suggestions that led to this illustration.

[11] Walter A. Hendricks, Chapter 5 in the book *Agricultural Estimating and Reporting Service* (Miscellaneous Publications No. 703, Bureau of Agricultural Economics, Washington, 1949); pages 31–35 in particular.

propriate, or it may be that some other scale will give more accurate results with convenience.

For an estimate of a^* by extrapolation, we may look upon recalls as necessary to provide the required coordinates of points by which to make the extrapolation, and not merely to provide additional returns to add to the initial attempt.

For this new type of estimate, the standard error would not be calculated in the usual way (Equation 24), but as the standard error of the intercept on the scale along which we read, by extrapolation, the estimate of a^*. New theory will be required for the optimum allocation of effort amongst the various recalls, and for effecting the extrapolation; also for calculating its standard error. It may turn out, for example, that unless one can achieve extremely high initial response, approaching 90%, there may be little point in expending funds to build it up. It is possible that theory beyond the scope of this paper may lead to efficiency and reliability far beyond those attained in practice today.

SOME REMARKS ON CLASS 0

We must face the fact that our survey can at best only provide estimates for Classes 1–6, although it can also give us the proportion p_0 and some of the characteristics of Class 0. The administrative decisions that the survey was expected to help may nevertheless involve Class 0 along with the others. In a marketing study, for example, the people in this class may be heavy purchasers of the very commodity that forms the subject of the survey. They may in part be people who travel much, and who may thus be important to a railway, an air line, a manufacturer of automobiles, a hotel, and to others. They may be people in high income groups. It may therefore be important to learn how much we are missing by not bringing them into the survey.

Unfortunately, it is impossible to learn this magnitude from the survey itself. The only possible approach seems to be from outside sources, such as through statistics on the total movement of a particular product from wholesale into retail stores. It is possible in many cases to gather outside evidence by which to evaluate approximately the magnitude of a_0 (the mean in Class 0), or rather of the total a_0p_0 in Class 0, for some of the important characteristics that affect the decisions or relate to them. The next step is to ascribe upper and lower bounds to the possible magnitude of a_0p_0, and thus to infer the possible effects of Class 0 on the uses and limitations of the data.[12]

[12] This suggestion came from Professor Philip M Hauser in an informal conversation in regard to this research.

The difficulty with Class 0 is not peculiarly a sampling problem, as Class 0 appears in complete counts as well as in samples—in fact, it is undoubtedly bigger in complete counts than in samples.

ACKNOWLEDGMENTS

This research could not have been completed without the generous assistance of a number of friends. In particular, my wife, Lola S. Deming, carried out the major portion of the calculations, and took charge of the preparation of the final drafts of the manuscript. Mr. J. E. Ball of the Monroe Calculating Machine Company lent me a machine for a long period while my own machine was under repair. Dr. Benjamin J. Tepping found a fundamental flaw in an earlier draft, and rendered indispensable assistance in the derivations of Equations 10 and 11. Messrs. Lester Frankel and Robert Weller of the Alfred Politz Research organization rendered prompt and helpful assistance on a number of occasions. My colleagues, Morris H. Hansen, William N. Hurwitz, and George Kuznets, kindly offered suggestions toward clarification.