

to 30, and $n = 30 \times 20 = 600$. If we have some idea of the proportion p in advance, we should use it in the formula, but if not it is best to use $p = 0.5$ as a guess since this maximizes s and hence errs on the safe side. The value of roh is hardest to estimate, but is likely to be high, with more variation in such an item between communities than within each community, so we may take $roh = 0.20$. Using the formula (3) we obtain a design effect of

$$D = 1 + (29 \times 0.20) = 6.8$$

and from (1) the estimate of the standard error is

$$s = \sqrt{[0.5 \times 0.5 \times 6.8 / 600]} = 0.05$$

or 5%. This indicates that with such a sample size we can be 95% certain that the true proportion of households with latrines will lie within $\pm 10\%$ (2 standard errors) of our estimate. Whether or not this precision is adequate depends on the purpose of our survey. If the design effect had been ignored, we would have predicted a standard error of

$$s = \sqrt{[0.5 \times 0.5 / 600]} = 0.02,$$

encouraging us to believe that our survey would give much more precise results than would actually be the case.

Suppose that in the same survey we also wished to estimate the proportion of children aged 12-23 months who had been adequately vaccinated by their first birthday. If we could assume that such children are found in about one-quarter of all households, then we would expect to get about 7 responses from each cluster, and we would take this as the value of b . The values of n would be $7 \times 20 = 140$. We might take the value of roh to be 0.10 and following the above calculations would obtain $D = 1.6$ and $s = 5.3\%$, giving a 95% confidence interval of about 11%. Ignoring D would have led us to underestimate the width of the confidence interval as 8%.

Estimating sample size

If the investigator knows that a certain precision is required from the survey, then the necessary sample size may be calculated. Usually it will be a matter of deciding how many cluster samples of a given size b will be necessary. The design effect D should be calculated from (3) as before, and then the number of clusters necessary is given by c where

$$c = \frac{p(1 - p)D}{s^2b}. \quad (4)$$

For example if p is expected to be around 20% for some measure of disease prevalence, for which we expect roh to be about 0.02, and suppose that we wish to estimate p to within $\pm 5\%$. If we expect to have 20 responses from each cluster, then the value of D will be 1.38 (from (3)). For a confidence interval of $\pm 5\%$ we shall need $s = 0.025$, then from (4) we need $c = 18$ clusters.

If we had failed to take account of the design effect we would have estimated the sample size from equation (4) as 13 clusters. Using equation (1), we see that our result would then have had a predicted standard deviation of 0.029 and a confidence interval of $\pm 6\%$, a little less precise than desired. The small size of the loss of precision in this example is due

only to the small value of D . In many cases, D will be considerably larger, and the precision achieved considerably less than desired. In general, ignoring the design effect in estimating the sample size required will lead to confidence intervals which are wider than desired by a factor of \sqrt{D} .

Such calculations should be made for the most important items in the survey schedule. Ideally c should be chosen to be the largest value given by these calculations in order to satisfy all the requirements. If the sample sizes necessary for different items are grossly different (as may happen in a study which covers both disease prevalence and usage of health-care facilities), it may be advisable to just use a subsample for those questions requiring fewer responses. However, the increase in complexity of the instructions given to interviewers means that this should be used with caution.

One should note that if the prevalence of an item under consideration is expected to be quite low, for example HIV seropositivity which may in some countries be around 2%, then it is not sensible to design a survey to achieve an absolute precision of 5%. In such a case the standard error desired needs to be considered relative to the expected prevalence rate, and would be much smaller, say 0.5% in absolute terms.

If the survey has been stratified (see Section 6) then each stratum should be considered as a separate survey, and sample-size calculations performed for each one to give the precision necessary for that stratum. The precision of the overall national estimate will then be somewhat better than that for any single stratum.

If the survey is one of a series, and the purpose is to estimate the change in some measure since the previous survey, then one needs to estimate the standard error of the change. This will be larger than the standard error of the new estimate of the measure, because of the imprecision of the estimate of the measure from the previous survey. To allow for this, the sample size may need to be double that calculated by the usual methods.

5. Analysis of data

This section describes the methods used to provide estimates of proportions or rates, together with standard errors of those estimates so that confidence intervals can be calculated. A mean value may also be estimated in the same way. We also describe how to calculate D and roh . The methods described below can be carried out on a simple calculator having a square-root key, and the use of a spreadsheet is illustrated in the Annex. The calculations in this and earlier sections may also be programmed easily on a computer using a spreadsheet package, as shown by Frerichs (11).

Estimation of a proportion

Suppose that a number of households have been selected in each of c communities with a view to estimating (by examining their record cards) what proportion of children aged 12-17 months were fully vaccinated on their first birthday. Suppose that in the i^{th} community ($i=1, \dots, c$) these were x_i children whose record cards were examined, and that y_i of these were fully vaccinated as defined by the study.

Then the proportion of children in the i^{th} community who were fully vaccinated will be given by

$$p_i = y_i/x_i.$$

In the survey population as a whole the proportion who are fully vaccinated will be estimated by

$$p = \sum y_i / \sum x_i \quad (5)$$

i.e. the total number of children vaccinated divided by the total number of children whose cards were examined. This is the straightforward ratio of the sample totals. Note that it is not the same as the average of the p_i 's, which would be incorrect since it does not take account of the variation in the x_i 's.

The standard error, s , of p is obtained from the formula

$$s = [c/\sum x_i] \sqrt{\{[\sum y_i^2 - 2p\sum x_i y_i + p^2\sum x_i^2]/[c(c-1)]\}}. \quad (6)$$

A spreadsheet for calculation of s is given in the *Annex*, with an example of its use. This formula is more complex than the formula (2) usually used by standard computer packages in that it takes account of (i) the clustering of the sample and (ii) the variability between clusters of the denominator x_i . This value, (the number of record cards examined in the i^{th} community) will have been unknown before the survey began and would probably be different if a different sample of households were taken from the same community. Failure to take account of these factors would lead to underestimation of s , and consequent overconfidence in the precision of the results (see *Annex* for an example). In many cases x_i will not vary much between communities, for example when x_i is the number of households selected, and then the simpler formula

$$s = \sqrt{\{\sum (p_i - p)^2 / [c(c-1)]\}} \quad (7)$$

may be used instead of (6).

Estimation of means

At times one will collect data on values which are not simply "yes/no" attributes of the household or person, but counts or other measurable quantities, for example "number of children ever born" or "number of rooms". In this case one may wish to estimate the mean value over the population, for example the mean number of children ever born (although of course one may also estimate a proportion, for example the proportion of women who have given birth to more than 3 children). Estimation of the mean and its standard error are carried out in exactly the same way as for a proportion (Section 5) except that y_i will now be equal to the sum of the numbers of children ever born to all of the x_i mothers interviewed in the i^{th} community.

Weighted analysis

In many situations there will be a need to weight the observations to allow for different probabilities of selection or different levels of non-response. For example suppose clusters were chosen with PPS as in Section 3, and it was intended to visit 25 households in each one, but because of staff illness it was only possible to visit 16 households in one of the clusters. If this fact is ignored, it will lead to that cluster being underrepresented in the calculation of the proportion p and its standard error. The solution is to weight the responses from this community by multiplying them up by 25/16. In more

general terms, this means replacing x_i and y_i each time they occur in formulae (5) and (6) with $w_i x_i$ and $w_i y_i$, giving the more general formulae

$$p = \sum w_i y_i / \sum w_i x_i$$

and

$$s = [c/\sum w_i x_i] \sqrt{\{[\sum w_i^2 y_i^2 - 2p\sum w_i^2 x_i y_i + p^2\sum w_i^2 x_i^2]/[c(c-1)]\}},$$

where w_i is the weight attached to the i^{th} cluster. An unweighted cluster has $w_i = 1$.

If clusters are sampled with probability proportional to size and x_i represents the number of BSUs (households) selected, then the proportion is estimated by \bar{p} , the average of the p_i 's, and we can use formula (7) for its standard error with p replaced by \bar{p} . In other cases the approximate formula (7) ignores the size of the cluster and should not be used if weighting is necessary.

Weighting may also be used to allow for clusters not being selected with probability proportional to size, for example when current size was not known at the time of their selection and they were selected with simple random sampling (or with probability proportional to a poor or very out-of-date measure of size). In this case the weight will be proportional to the actual population of the cluster (or the ratio of this to its old estimate).

Estimation of design effect

The results of any survey may be used to estimate design effects, for use in the same or future surveys. The design effect is estimated by

$$D = \frac{s^2 \text{ from equation (6) or (7)}}{s^2 \text{ from equation (2)}}$$

The rate of homogeneity, roh , may then be estimated as

$$(D - 1)/(b - 1)$$

where b is as defined earlier. An example is given in the *Annex*.

Imputation of standard errors

In a large survey it may not be feasible to use the correct formulae (6) or (7) to estimate the standard error of every variable. In such a case one may calculate exact standard errors for a few variables of each type (socioeconomic, health status, etc.). Dividing each standard error by the corresponding binomial value (2) gives a new estimate of the design factor (the square root of the design effect \sqrt{D}). For the remaining variables of the survey the simple formula (2) as given by calculator or standard software can be used, and just multiplied by the most appropriate value of \sqrt{D} obtained for variables of similar type.

6. Extensions

The previous sections describe cluster-sampling procedures in a simple context: a sample of communities is selected from the whole region under consideration and a sample of households is visited in each selected community. Such a sampling scheme will be inadequate if the region is very large or if separate estimates are needed for different

geographical areas. In this section we show how the techniques described above can be extended to allow for multistage sampling and stratification.

Multistage sampling

In a large region or country where an overall estimate is required, it will usually be sensible to select the sample of communities in at least two stages. For example, if the country is split into a number of administrative districts one would take a sample of districts by the systematic PPS method described in Section 3 (i.e. by making a list with cumulative population sizes). Within each selected district, communities would be selected, again by the systematic PPS method. The same number of communities must be selected in each district. If some districts are very small it may be sensible to combine them. Households would be selected in the usual way, with again the same number selected in each community.

With the systematic PPS method described here it is possible that the same district may be selected twice. This will happen if the population of the district is larger than the sampling interval. In this case two independent samples of communities should be selected from this district.

Decisions on the sample size will be made exactly as in Section 4, except that b will now be the expected number of responses per district and c will be the number of districts in the sample. The value of roh is now an indicator of the ratio of between-district variances to within-district variances. In theory, this requires an estimate of roh from a survey of similar multistage design. In practice, such estimates are not available, and the best one can do is probably to use the values given in Section 4 as guidelines, and bear in mind that they will be overestimates, as the value of roh is likely to decline slowly with the size of the primary cluster used.

The analysis will follow exactly the same pattern as in Section 5 except that x_i and y_i now refer to the number of responses and the number of positive responses respectively in the i^{th} district, summed over all communities selected in that district.

The method of sampling described here may be extended in exactly the same way to more stages if required.

Stratification

It may be necessary to obtain separate estimates for, say, the urban and rural sectors of the population, or for different provinces or ecological zones. Each province (etc.) will be a stratum, and a sample should be selected independently from each stratum. The sample size and structure for each stratum should be chosen with the conditions and needs of that stratum in mind, as if a separate survey were being carried out in that stratum alone. The samples may be of a different type and/or size for each stratum.

An estimate for each stratum may be calculated together with its standard error by treating each stratum as a separate survey. A stratified estimate for the whole country may then be calculated by weighting the stratum estimates by the stratum populations. For example, suppose there are three strata and the estimates from them are p_1 , p_2 and p_3 , with

standard errors s_1 , s_2 and s_3 respectively. Then the estimate for the whole country would be

$$p = V_1p_1 + V_2p_2 + V_3p_3$$

with standard error

$$s = \sqrt{[V_1^2s_1^2 + V_2^2s_2^2 + V_3^2s_3^2]}$$

where V_1 is the proportion of the country's population which belongs to stratum 1, and so on ($V_1 + V_2 + V_3 = 1$). The standard error s for the national estimate will be somewhat less than the standard errors for the individual strata.

Implicit stratification

Stratification usually leads to a small reduction in the standard error of the overall estimate p , compared to the error that would have been obtained if the survey had not been stratified. Another way of obtaining such a reduction is by *implicit stratification*. This is simply carried out at the time of selection of communities (or districts) by ensuring that the list of communities from which the systematic sample is to be taken is ordered by some measure which is correlated with the main purpose of the survey. For example, in a survey of the utilization of mother-and-child health facilities, there may have been a previous study carried out some years ago on the same subject, or there may be other knowledge available which indicates which communities may be expected to have high levels of utilization and which communities low levels. If not, one may be able to guess that those communities which are further from the regional capital, or which cover a more widely-scattered population, will have lower levels of utilization than others. Whatever the measure chosen, if the communities can be listed in approximate order from a high to a low level of expected utilization, then the sample selected will contain communities with a spread of utilization levels, and the estimated proportion p will be more precise. The standard error will be reduced, and its estimate s given by (6) will be somewhat of an overestimate (15). The improvement in precision cannot be quantified adequately to allow its use in sample-size calculations.

7. Conclusion

A simplified approach to survey design has been presented, with no attempt to cover all possible types of estimation. We have rather aimed to provide a set of guidelines which will enable the practitioner to plan a survey in a way which will give a reasonably representative sample, without any great bias, and of a suitable size to give adequate precision without wasting resources. The values given for the rate of homogeneity have of necessity been approximate, but variability between surveys and between variables is such that precise advice is impossible. The methods of analysis presented here offer an improvement on the common practice of assuming that the data came from a simple random sample and using the standard errors given by a calculator or standard computer package.

Acknowledgements

This article had its origins in work carried out for WHO/SHS on sampling in primary health care reviews. We are grateful to S. Lwanga, S.A. Sapirie, J. L. Tulloch and J. E. Dowd of the World Health Organization for their contributions, and to many colleagues for helpful comments.

SUMMARY

General guidelines are presented for the use of cluster-sample surveys for health surveys in developing countries. The emphasis is on methods which can be used by practitioners with little statistical expertise and no background in sampling. A simple self-weighting design is used, based on that used by the World Health Organization's Expanded Programme on Immunization (EPI). Topics covered include sample design, methods of random selection of areas and households, sample-size calculation

and the estimation of proportions, ratios and means with standard errors appropriate to the design. Extensions are discussed, including stratification and multiple stages of selection. Particular attention is paid to allowing for the structure of the survey in estimating sample size, using the design effect and the rate of homogeneity. Guidance is given on possible values for these parameters. A spreadsheet is included for the calculation of standard errors.

RÉSUMÉ

Méthodes générales simplifiées pour les enquêtes sanitaires utilisant le sondage par grappes dans les pays en développement

Cet article présente des directives générales concernant l'exécution d'enquêtes sanitaires utilisant le sondage par grappes dans les pays en développement. L'accent y est mis sur des méthodes utilisables par des praticiens peu spécialisés en statistique et sans formation de base en matière de sondages. Ces méthodes font appel à un plan comportant un dispositif d'auto-évaluation simple, inspiré de celui qui est utilisé par le Programme élargi de vaccination (PEV) de l'Organisation mondiale de la Santé. Les sujets traités couvrent le plan de sondage, les méthodes de sélection aléatoires des zones et des

ménages, le calcul de la taille des échantillons et l'estimation des proportions, des taux et des moyennes, avec les erreurs types appropriées au plan. L'article traite aussi de questions telles que la stratification et les différentes étapes de la sélection. On insiste sur l'importance de tenir compte de la structure des enquêtes pour estimer la taille des échantillons, en utilisant l'«effet de plan» et le taux d'homogénéité. Des conseils sont donnés sur les valeurs qu'il serait possible d'attribuer à ces paramètres. Une feuille de calcul pour les erreurs types est jointe.

ANNEX

Estimating the standard error of a ratio and its design effect

The use of a simple spreadsheet for the calculation of an estimate and its standard error using the precise formula (6) is demonstrated using the following example. The use of the approximate formula (7) for the standard error is also shown, and the design effect is calculated. The sample size is much smaller than those encountered in practice but all the important steps in the calculation are demonstrated.

Six communities are selected using the systematic PPS procedure. Twenty households are chosen in each community in order to estimate, for the population, the proportion of recently-pregnant mothers who have received postnatal care.

The data are:

Community	Number of recently-pregnant women	Number receiving postnatal care
1	2	2
2	7	5
3	4	3
4	6	3
5	4	1
6	3	0

The spreadsheet is constructed as follows:

	y_i	x_i	y_i^2	x_i^2	$x_i y_i$	p_i
	2	2	4	4	4	1.00
	5	7	25	49	35	0.70
	3	4	9	16	12	0.75
	3	6	9	36	18	0.50
	1	4	1	16	4	0.25
	0	3	0	9	0	0.00
Total	A=14	B=26	C=48	E=130	F=73	

Here $c=6$ is the number of communities; y_i is the number of recently-pregnant mothers in the i^{th} community who have received postnatal care; x_i is the number of recently-pregnant mothers in the sample from the i^{th} community.

The estimated proportion is

$$p = A/B = 0.5385.$$

The standard error s , as given by (6), is calculated as follows:

New quantity	Calculated as	Value
p^2	$p \times p$	0.2900
G	$2 \times p \times F$	78.621
H	$p^2 \times E$	37.70
J	$C - G + H$	7.079
K	$J / [c \times (c-1)]$	0.2360
L	\sqrt{K}	0.4858
s	$c \times L / B$	0.1121

The 95% confidence interval for the true proportion is $0.5385 \pm (2 \times 0.1121)$, i.e. 0.314 to 0.763.

The approximate formula (7) gives $s = 0.1482$. The difference between this figure and that given above arises because the x_i 's are very variable.

The standard error assuming a simple random sample is given by (2) as

$$s_{srs} = \sqrt{\{(0.5385) \times (1-0.5385)/26\}} = 0.0978,$$

thus ignoring the design of the study would have led us to assign our estimate a confidence interval from 0.343 to 0.734, which is 13% narrower than the correct value.

The design effect is estimated as

$$D = s^2/s_{srs}^2 = (0.1121)^2/(0.0978)^2 = 1.31.$$

Since $b = \sum x_i/6 = 4.333$, roh may be estimated in this case by $(D - 1)/(b - 1) = 0.093$.

REFERENCES — RÉFÉRENCES

- HENDERSON, R. H. & SUNDARESAN, T. Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method. *Bulletin of the World Health Organization*, **60** (2): 253-260 (1982).
HENDERSON, R. H. & SUNDARESAN, T. Sondage par grappe pour évaluer la couverture vaccinale: examen d'une expérience sur une méthode simplifiée de sondage [résumé]. *Bulletin de l'Organisation mondiale de la Santé*, **60** (2): 259-260 (1982).
- LEMESHOW, S. & ROBINSON D. Surveys to measure programme coverage and impact: a review of the methodology used by the Expanded Programme on Immunization. *World health statistics quarterly*, **38** (1): 65-75 (1985).
LEMESHOW, S. & ROBINSON, D. Enquêtes sur la couverture et l'impact des programmes: méthodes quantitatives utilisées par le Programme élargi de vaccination. *Rapport trimestriel de statistiques sanitaires mondiales*, **38** (1): 65-75 (1985).
- LEMESHOW, S. ET AL. A computer simulation of the EPI survey strategy. *International journal of epidemiology*, **14**: 473-481 (1985).
- ROTHENBERG, R. B. ET AL. Observations on the application of EPI cluster survey methods for estimating disease incidence. *Bulletin of the World Health Organization*, **63** (1): 93-99 (1985).
ROTHENBERG, R. B. ET AL. Utilisation des méthodes de sondage par grappes du PEV pour estimer l'incidence d'une maladie: observations [résumé]. *Bulletin de l'Organisation mondiale de la Santé*, **63** (1): 97-98 (1985).
- COCHRAN, W. G. *Sampling techniques*. Chichester, John Wiley & Sons, 1977.
- KISH, L. *Survey sampling*. London, John Wiley & Sons, 1965.
- EL BINDARI-HAMMAD, A. & SMITH, D. L. *Primary health care reviews, guidelines and methods*. Geneva, World Health Organization, 1989.
- VERMA, V. *The estimation and presentation of sampling errors*. London, World Fertility Survey, 1982. (Technical Bulletin No. 11).
- LEMESHOW, S. ET AL. *Adequacy of sample size in health studies*. Chichester, John Wiley & Sons, 1990.
- FREERICH, R. R. & TAR TAR, K. Computer-assisted rapid surveys in developing countries. *Public health reports*, **104**: 14-23 (1989).
- FREERICH, R. R. Simple analytic procedures for rapid microcomputer-assisted cluster surveys in developing countries. *Public health reports*, **104**: 24-35 (1989).
- CASLEY, D. J. & LURY, D. A. *Data collection in developing countries*. Oxford, Oxford University Press, 1981.
- ABRAMSON, J. H. *Survey methods in community medicine*. Edinburgh, Churchill Livingstone, 1984. (Third edition).
- KISH, L. ET AL. *Sampling errors for fertility surveys*. London, World Fertility Survey, 1976. (Occasional paper No. 17).
- KISH, L. & HESS, I. On variances of ratios and their differences in multistage samples. *Journal of the American Statistical Association*, **54**: 416-446 (1959).