CHAPTER V

DISCUSSION AND CONCLUSION

5.1 Summary

The sampling design for Media Index Survey (MIS) is a stratified 2-stage cluster sampling. The representative sample of 8738 households in Peninsular Malaysia is self-weighting. The sample selection was performed utilizing area sampling frame. The primary sampling units (PSUs) are the enumeration blocks of approximately 200 to 240 dwelling units. The overall sample for the MIS has an average design effect (*deft*) of 1.77, and an average rate of homogeneity of 0.12. Therefore, on average, the variances in MIS will increase by 3.13 times as compared to simple random sampling design.

The results show that the urban-rural differentials in the mean design effects over all variables are rather small. This can be attributed to the similar sampling design and cluster size used across all strata, which are rather homogeneous. However, the design effects by Variable Category vary between and within geographical domains.

Differentials in design effects across ethnicity are more discernible but less so across age groups, because of the fact that population clustering is more obvious by ethnic group than age group. Differentials in design effects across gender are rather small for variables measuring household information, but the gender differentials are relatively larger for variables measuring individual information.

Table 5.1 illustrates few design effect values obtained using stratified cluster sampling. The total *deft* values ranging from 1.47 to 1.91 with the background and 'life time' variables skewed towards the upper end of the range. In view of the large

number of sample base and variables included in the calculation, it is fairly certain that the average design effect values for surveys with similar design are likely to fall within the range mentioned above. Compared to most other surveys using complex sampling designs, the design effects of the MIS are within acceptable range (see Table 5.1)

Table 5.1: Design Effect Values for Complex Sampling Designs

	Urban	Rural	Total
Media Index Survey	1.82	1.69	1.77
South Korea Fertility Survey (1973)*	-	-	1.47
Mean *deft* over 39 variables			
Malaysia Family Planning Acceptors Survey (1969)*	1.01	1.92	1.61
Mean *deft* over 29 variables			
Peru Fertility and KAP Survey*	-	-	1.65
Mean *deft* over 29 variables			
DHS Surveys**			
Mean *deft* of background or 'life time' variables over 48 counties			
a. Illiterate	1.74	1.90	1.91
b. Ever used of contraception	1.57	1.74	1.74
c. Ideal family size	1.55	1.66	1.71
d. Age marriage	1.47	1.44	1.50

* Kish et al., 1976

** Verma et al., 1996

The coefficients of variation for all the variables presented in Table 4.9 are in he range of about 5% and below. The standard errors and hence the 95% confidence intervals are not too large to cause concerns on the accuracy of the estimates. Therefore, the stratified 2-stage cluster sampling design is adequate and suitable for Media Index Survey.

The results in Table 4.9 are based on large sample base of 8738 interviews. However, the choice of sample size in most of the national studies especially in market research is around 2000 interviews (Aaker, David A. et al., 1995). Then, what is the impact on the coefficients of variation of sample estimates for MIS sampling design if the sample size is reduced?

The coefficients of variation for selected variables based on sub-sample of field quarters of quarter 3, 1997 and quarter 1, 1998 were computed. The sample size for this sub-sample is approximately half the size (4435) of the original MIS sample. Table 5.2 shows the coefficients of variation for total sample and sub-sample. The coefficients of variation for sub-sample increased by about 40% for each of the selected variables in sub-sample.

The outcome is still quite acceptable as most of the coefficients of variation are maintained at below 5% except for variables DURA35 (air conditioner ownership) and P1MBUY18 (bought baby diapers in the past 1 month). Both of the variables have the estimate values below 0.2. In other words, these 2 products have a penetration rate below 20%.

This further supports the appropriateness of the stratified 2-stage cluster ıpling design adopted in MIS. However, if one wishes to reduce the sample size, n one must ensure that the coefficients of variation of a characteristic fall within acceptable range to fulfil the research objectives.

Table 5.2: Sampling Errors of Selected Variables - Total Sample and Sub-Sample, MIS

Variable	Design effect	Value	Std error	Relative error	Value	Std error	Relative error
		Total Sample (n=8738)			**Sub-Sample (n=4435)**		
	deft	r	se	se/r	r	se	se/r
URA12	1.7328	0.7249	0.0083	0.0114	0.7211	0.0117	0.0162
URA29	1.6049	0.8816	0.0055	0.0063	0.8778	0.0079	0.0090
URA35	2.0780	0.1441	0.0078	0.0542	0.1409	0.0109	0.0770
IAABV15	1.4968	3.0409	0.0236	0.0078	3.0570	0.0336	0.0110
IMBUY15	1.9811	0.4034	0.0104	0.0258	0.4070	0.0146	0.0359
'IMBUY18	1.4134	0.1192	0.0049	0.0411	0.1238	0.0070	0.0565
TVSET	1.5941	1.1214	0.0081	0.0073	1.1389	0.0119	0.0104
ACTVT1	1.5633	0.2470	0.0072	0.0292	0.2586	0.0103	0.0397
BANK13	1.6909	0.4261	0.0089	0.0210	0.4352	0.0126	0.0289
FREQ1	2.3625	4.3748	0.0779	0.0178	4.3594	0.1082	0.0248
LITER3	1.7671	0.7941	0.0076	0.0096	0.7955	0.0107	0.0135
NPYTDA	1.8033	0.5699	0.0096	0.0168	0.5691	0.0134	0.0236

The following section will examine the cost implications if one wishes to improve the efficiency by reducing the average *deft* value by 10%, which is from 1.77 to 1.59. With the formulation $roh = (deft^2 - 1)/(\bar{b} - 1)$, the average number of elements per cluster ($\bar{b}$) can be calculated by substituting *roh*=0.12 and $deft^2 = 2.53$ in the above equation. The $\bar{b}$ is now reduced from 19.9 to 13.8 elements per cluster. Then, the number of PSUs will eventually increase from 439 to 633 with respect to the same sample size. Assuming that the travelling cost increases proportionately with the number of PSUs, a 10% improvement of design effect (*deft*) will incur about 44% increases in travelling cost to conduct the similar survey.

Conversely, if one wishes to save cost and accept a larger *deft* value, say, by from 1.77 to 1.95, the number of PSUs will be reduced from 439 to 358 with ect to the same sample size. Then, the travelling cost of conducting the similar ɾey will be reduced by 23%, for a 10% increase in design effect (*deft*).

Conclusion

The stratified 2-stage cluster sampling adopted in MIS utilized area sampling ıme. This sampling design is commonly used for nationwide sample surveys. ıta collection method employed is mostly face-to-face personal interviewing. To ıte, personal interviewing is still the most widely used data collection method ɜpecially in general public sample survey. It has been considered as the primary ata collection method.

This conclusion was based on a number of factors. Face-to-face Personal ınterviewing has several advantages over other methods of data collection. The method allows the use of special questionnaire techniques, show product concepts and show-cards. Besides, this is the only feasible way to conduct long, in-depth and in-homes product tests. However, there are 2 major drawbacks in face-to-face personal interview. This method is always more costly and need longer data collection period.

The rapid advancement in information technology is changing the process of data collection. Computers and communication technology are revolutionizing the way surveys are conducted. The telephone interviews have gradually become the dominant data collection method especially in areas where the telephone penetration rate is high.

Random-digit dialing and systematic random-digit dialing are approaches that have been developed to obtain representative telephone numbers for telephone interviews. This method gives all telephone numbers an approximately equal chance of being dialed. The development of Computer-Assisted Telephone Interview (CATI) further enhances the capability of telephone interview. Telephone interview has become more popular, as compared to face-to-face interview. Telephone interview is less costly, needs shorter data collection period and is able to have access to population of disperse localities. Besides, the random-digit dialing sampling is also able to provide representative samples. This method, however, does not allow the use of visual aids, and is unsuitable for long and complicated questionnaire.

Electronic mail or Internet surveys are gradually becoming another popular data collection method even though it is only applicable to specific target group now.

Technology development is changing the survey method. The use of telephone interview and Internet survey has led to a move away from using area sampling frame to perform the sample selection. For example, in telephone interview, the random digit dialing method is able to provide an element sampling of households with telephone. These developments have changed the use of complex sampling designs back to simple random sampling.

Nevertheless, the survey method used is also dependent on the target population, which means that face-to-face personal interview will still be the dominant mode of data collection method especially in the nationwide general public surveys. The survey researchers have to ensure that every target population have a known non-zero probability of inclusion in the surveys. For example, data

set from MIS shows that only 57% of households in rural areas reported owning telephone. This indicates that the telephone ownership and internet access rate are not yet universal enough for conducting surveys using these communication modes.

Therefore, complex sample designs are unavoidable in sample surveys at least in large-scale national sample surveys. Sampling errors computation should take into account the sampling design in order to provide accurate statistical inferences. The study of design effects of Media Index Survey in Peninsular Malaysia will enable us to provide a general guideline on the design effects patterns of stratified 2-stage cluster sample.