

Single Stage Cluster Sampling in Prevalence-Incidence Surveys: Some Issues Suggested by the Shanghai Survey of Alzheimer's Disease and Dementia

ZHISEN XIA, PAUL S. LEVY, ELENA S.H. YU, ZHENGYU WANG,
and MINGYUAN ZHANG¹

ABSTRACT

The scenario considered here is that of a sample survey having the following two major objectives: (1) identification for future follow up studies of n^* subjects in each of H subdomains, and (2) estimation as of this time of conduct of the survey of the level of some characteristic in each of these subdomains. An additional constraint imposed here is that the sample design is restricted to single stage cluster sampling. A variation of single stage cluster sampling called *telescopic single stage cluster sampling* (TSSCS) had been proposed in an earlier paper (Levy *et al.* 1989) as a cost effective method of identifying n^* individuals in each sub domain and, in this article, we investigate the statistical properties of TSSCS in crosssectional estimation of the level of a population characteristic. In particular, TSSCS is compared to ordinary single stage cluster sampling (OSSCS) with respect to the reliability of estimates at fixed cost. Motivation for this investigation comes from problems faced during the statistical design of the Shanghai Survey of Alzheimer's Disease and Dementia (SSADD), an epidemiological study of the prevalence and incidence of Alzheimer's disease and dementia.

KEY WORDS: Single stage cluster sampling; Prevalence estimation; Telescopic single stage cluster sampling; Alzheimer's disease; Dementia.

1. BACKGROUND AND INTRODUCTION

Many studies have both a crosssectional component in which the levels of quantitative variables or prevalences of dichotomous variables are estimated by means of a sample survey, and a longitudinal component in which a cohort of individuals is identified by means of the same sample survey and followed over a defined period for subsequent events. This type of study is especially common in the field of epidemiology in which estimates of the prevalence of a disease or condition are required both for the study population as a whole as well as for defined subgroups of it, and a sufficient number of individuals initially free of the disease or condition need to be identified within each of the defined subgroups for future estimation of the incidence of the disease or condition (*cf.* Kannel 1966).

Design of a cost efficient sampling plan for such studies poses a challenge since sufficient numbers of individuals within each domain must be selected, often under some type of cluster sampling scheme, to ensure reliable estimation of both the prevalence and incidences discussed above. In this report, which has been motivated by a recent study conducted in China, we discuss these issues of sample design under a particular type of cluster sampling (single stage cluster sampling).

¹ Zhisen Xia, Paul S. Levy and Zhengyu Wang, School of Public Health, University of Illinois at Chicago, P.O. Box 6998, Chicago IL. 60680, U.S.A., Elena S.H. Yu, School of Public Health, San Diego State University, San Diego, California, U.S.A., Mingyuan Zhang, Shanghai Institute of Mental Health, Shanghai, Peoples Republic of China.

2. STATISTICAL FORMULATION

Let us suppose that a population consists of N individuals divided into H mutually exclusive subdomains, each containing N_h individuals ($h = 1, \dots, H$). Suppose further that the population is grouped into M clusters which will comprise the sampling units for the survey. Let us assume that sampling of the clusters will be according to ordinary single-stage cluster sampling (*i.e.*, simple random sampling of clusters followed by selection of all individuals within each sample cluster.)

If we wish to identify with $100 \times (1 - \alpha)\%$ confidence at least n_h^* individuals in a particular domain, h , then the following number, m'_h , of clusters must be selected (*cf.* Levy *et al.* 1989):

$$m'_h = \left[A_h + \left(A_h^2 + \frac{n_h^*}{\bar{N}_h} \right)^{1/2} \right]^2, \quad (1)$$

where

N_{hi} = the number of individuals in domain h , cluster i , ($i = 1, \dots, M$),

$$\bar{N}_h = \sum_{i=1}^M N_{hi}/M,$$

$$V_{N_h} = \sigma_{N_h}/\bar{N}_h,$$

$$\sigma_{N_h}^2 = \sum_{i=1}^M (N_{hi} - \bar{N}_h)^2 / (M - 1),$$

z_α = the 100α 'th percentile of the normal distribution

and

$$A_h = |z_\alpha| \times V_{N_h}/2.$$

The above assumes that the N_{hi} are normally distributed over the M clusters. Also, the number, n_h^* , of individuals needed in domain h is based on statistical considerations relevant to the longitudinal component of the study. For example, it could be based on the expected occurrence rate of the event of interest in the follow up period and the precision required for the estimate of this occurrence rate.

If one also wishes to estimate with $100 \times (1 - \alpha)\%$ confidence the total or mean level of some variable \mathcal{H} to within $100 \times \epsilon\%$ of its true value for each domain, h , then one would require sampling of the following number, M''_h , of clusters in domain h ;

$$m''_h = \frac{z_{1-\alpha/2}^2 M V_{hx}^2}{z_{1-\alpha/2}^2 V_{hx}^2 + (M - 1)\epsilon^2}, \quad (2)$$

where,

X_{hij} = the level of variable \mathcal{X} for individual j within domain h of cluster i
 $(j = 1, \dots, N_{hi}; i = 1, \dots, M),$

$$X_{hi} = \sum_{j=1}^{N_{hi}} X_{hij},$$

$$\bar{X}_h = \sum_{i=1}^M X_{hi}/M,$$

$$\sigma_{hx}^2 = \sum_{i=1}^M (X_{hi} - \bar{X}_h)^2/M,$$

and

$$V_{hx}^2 = \sigma_{hx}^2 / \bar{X}_h^2.$$

For both of the specifications stated above to be satisfied within each domain, it follows that we would require m_h clusters to be sampled where for $h = 1, \dots, H,$

$$m_h = \max(m'_h, m''_h). \quad (3)$$

Without loss of generality, we can relabel the domains in order of increasing required m_h (i.e., $m_1 \leq m_2 \leq \dots \leq m_H$).

Finally, in order for both of the specifications to be satisfied in each of the H domains under an ordinary single stage cluster sampling design, the number, m , of clusters required to be sampled would be m_H . We note again that in ordinary single stage cluster sampling, every individual in every sample cluster is sampled. Thus, while the specifications of sample size are met minimally in domain H , the domain requiring the largest number of sample clusters, they are more than met in the other domains: $1, \dots, H - 1$. This inclusion in domains other than H of more individuals than are actually required could result in a survey that has overly expensive field costs.

The alternative to ordinary single-stage cluster sampling that is generally used to avoid this needless expense would be a two-stage cluster sampling design with different second stage sampling fractions (i.e., over sampling) in each domain. Given, however, a scenario in which it is not feasible to subsample at all within clusters, a methodology called *single stage telescopic cluster sampling* (SSTCS) was proposed in an earlier publication (Levy *et al.* 1989) which allowed the *eligibility rule* (i.e., the rule that determines which individuals are eligible for inclusion in the sample) to vary over the sample clusters. In this design, the particular domains included in the sample would not be the same for each sample cluster. This earlier publication demonstrated the usefulness of single stage telescopic sampling in surveys which have as major objective the identification for future longitudinal follow up of a certain number of individuals in each of several domains. In this report, we will characterize the properties of estimates from this type of design and compare them to estimates from ordinary single-stage cluster sampling.

3. TELESCOPIC SINGLE-STAGE CLUSTER SAMPLING

3.1 Sampling of Clusters

As mentioned above, single-stage telescopic cluster sampling is proposed as a cost saving alternative to ordinary single-stage cluster sampling in situations where it is not feasible to sub-sample within sample clusters, and is performed as follows. If there are H mutually exclusive and exhaustive domains for which estimates are desired, and if m clusters are to be sampled, the m sample clusters are divided randomly into m_1^* type 1 clusters, m_2^* type 2 clusters, \dots , and m_H^* type H clusters having the following properties: A type h cluster ($h = 1, \dots, H$) as illustrated below has as eligible sample persons individuals in domains $h, h + 1, \dots, H$, but not in domains h' where $h' < h$.

Cluster Type	Domains Sampled			
	1	2	h	H
1	+	+	+	+
2	–	+	+	+
h	–	–	+	+
H	–	–	–	+

“+” = domain sampled “–” = domain not sampled.

The term *telescopic* was suggested by the appearance of the above diagram.

The number, m_h^* , of type h clusters is generally determined according to the following strategies: Suppose that under single-stage cluster sampling, a sample of m_h clusters as determined by relation (3) is required for domain h , ($h = 1, \dots, H$); and, again supposing that $m_1 \leq m_2 \leq \dots \leq m_H$, we would let:

$$m_1^* = m_1; \quad \text{and} \quad m_h^* = m_h - m_{h-1} \quad \text{for} \quad h = 2, \dots, H.$$

Clearly, this allocation results in a total of m_H sample clusters being selected, with elements in each domain, h , being sampled in m_h sample clusters, exactly the number of clusters required to achieve the specifications placed on the reliability of estimates and the identification of individuals for future follow up. As discussed above, if ordinary single-stage cluster sampling (OSSCS) were used, a sample of m_H clusters would be needed to meet specifications in domain H , but this would entail individuals in the other domains also being sampled in m_H clusters in excess of that needed to meet the stated specifications.

3.2 Characterization of Estimates

Let

$$\sigma_{h k x} = \sum_{i=1}^M (X_{hi} - \bar{X}_h)(X_{ki} - \bar{X}_k) / M,$$

$S_h = \{i_1, i_2, \dots, i_{m_h}\}$ = the set of sample clusters having eligible persons in domain h .

The following results can then be obtained from combinatorial theory.

1. The estimated total, x'_{tel} , under TSSCS of a population total X is given by

$$x'_{tel} = \sum_{h=1}^H x'_h, \quad (4)$$

where x'_h is given by

$$x'_h = (M/m_h) \sum_{i \in S_h} X_{hi}.$$

2. The mean, $E(x'_{tel})$, and variance, $\text{Var}(x'_{tel})$, of x'_{tel} are given by

$$E(x'_{tel}) = X, \quad (5)$$

$$\text{Var}(x'_{tel}) = \sum_{h=1}^H \frac{M^2}{m_h} \left(\frac{M - m_h}{M - 1} \right) \left(\sigma_{hx}^2 + 2 \sum_{k < h} \sigma_{h k x} \right). \quad (6)$$

These relationships follow in a straightforward way from combinatorial theory.

4. COST COMPARISONS BETWEEN OSSCS AND TSSCS

We can examine the comparative costs of OSSCS vs. TSSS by considering the following simple cost function that would be associated with OSSCS:

$$C_0 = C_1 m_H + C_2 m_H (\bar{N}_1 + \bar{N}_2 + \dots + \bar{N}_H) = m_H \left(C_1 + C_2 \sum_{h=1}^H \bar{N}_h \right), \quad (7)$$

where C_0 is the expected cost, C_1 is the cost component associated with clusters (*e.g.*, travel to and from cluster, procurement of the list of enumeration units in the cluster, preparation of materials for field work within the cluster, *etc.*) and C_2 is the cost component associated with listing units (primarily travel between listing units and interviewing). It should also be noted that the expression, $\sum_{h=1}^H \bar{N}_h$, is the average number of listing units per cluster. Again, throughout this discussion the listing units are the individuals themselves. The analogous expected cost, C_t , associated with telescopic sampling would then be given by:

$$C_t = C_1 m_H + C_2 (m_1 \bar{N}_1 + m_2 \bar{N}_2 + \dots + m_H \bar{N}_H) = m_H \left(C_1 + C_2 \sum_{h=1}^H \gamma_h \bar{N}_h \right), \quad (8)$$

where $\gamma_h = m_h/m_H$ (which is ≤ 1). Thus, the cost, C_t , associated with TSSCS is less than or equal to that associated with an OSSCS of the same number of clusters with the difference being equal to

$$C_2 m_H \sum_{h=1}^H (1 - \gamma_h) \bar{N}_h.$$

The most important comparison between the two sample designs, in many instances, would be that involving their performance at equivalent cost in estimating the overall level, X , of a characteristic, \mathcal{C} . An estimator, x'_{ord} , based on an OSSCS of m_H clusters (the number required to meet the specifications within each domain) would have variance given by:

$$\text{Var}(x'_{\text{ord}}) = \frac{M^2}{m_H} \left(\frac{M - m_H}{M - 1} \right) \sum_{h=1}^H \left(\sigma_{hx}^2 + 2 \sum_{k < h} \sigma_{h k x} \right). \quad (9)$$

This is not the usual form of the variance (*cf.* Levy and Lemeshow 1991, chapter 9), but is an algebraically equivalent form that can be compared directly with the variance of x'_{tel} based on a TSSCS design with m_H clusters sampled (equation (6)). The difference between these two variances is given by

$$\text{Var}(x'_{\text{tel}}) - \text{Var}(x'_{\text{ord}}) = \frac{M^3}{M - 1} \sum_{h=1}^H \left(\frac{m_H - m_h}{m_H m_h} \right) \sum_{h=1}^H \left(\sigma_{hx}^2 + 2 \sum_{k < h} \sigma_{h k x} \right), \quad (10)$$

which is greater than or equal to zero (0). This is not surprising since an OSSCS of m_H clusters will invariably result in a larger overall sample size than a TSSCS of the same number of clusters.

Although an OSSCS of m_H clusters will result in an estimator, x'_{ord} , which has a lower variance than the estimator, x'_{tel} , resulting from a TSSCS of the same number, m_H , of clusters, it does so at a higher cost. For this reason, it is more reasonable to compare x'_{tel} based on a sample of m_H clusters to x'_{ord} based on a sample of m^* clusters where m^* is the number of clusters that can be sampled from an OSSCS design at cost equivalent to that based on a TSSCS design having m_H sample clusters. From equations (7) and (8), it follows that m^* is given by:

$$m^* = m_H \left(\frac{1 + \frac{C_2}{C_1} \sum_{h=1}^H \gamma_h \bar{N}_h}{1 + \frac{C_2}{C_1} \sum_{h=1}^H \bar{N}_h} \right). \quad (11)$$

It should be noted that

$$(1) \quad m^* \leq m_H.$$

$$(2) \quad \text{As } C_2/C_1 \rightarrow \infty, \text{ then } m^* \rightarrow \bar{m}_w$$

$$\text{where,} \quad \bar{m}_w = \sum_{h=1}^H m_h \bar{N}_h \bigg/ \sum_{h=1}^H \bar{N}_h.$$

$$(3) \quad \text{As } C_2/C_1 \rightarrow 0, \text{ then } m^* \rightarrow m_H$$

and

$$(4) \quad m^* \text{ decreases monotonically with increase in } C_2/C_1 \text{ which implies that } \bar{m}_w \leq m^* \leq m_H.$$

From the above analysis, we note that at a cost equivalent to that of a TSSCS of m_H clusters, the variance of x'_{ord} (ignoring the finite population correction) will be inflated by at most a factor equal to m_H/\bar{m}_w over that which would have been obtained from an OSSCS of m_H clusters, where \bar{m}_w is a weighted mean of the m_h clusters required within each domain for the domain specific specifications to be met. The weights in this instance are the \bar{N}_h , which are the average number of individuals within each particular domain. It should be noted also that the reduction in effective sample size of an OSSCS equivalent in cost to a TSSCS increases with increase in C_2/C_1 , which is essentially the ratio of the cost of extracting information from sample individuals to that of preparing the sample clusters for the survey. This makes sense intuitively.

The issues discussed above are illustrated in the next section with data from the Shanghai Survey of Alzheimer's Disease and Dementia.

5. THE SHANGHAI SURVEY OF ALZHEIMER'S DISEASE AND DEMENTIA

The SSADD was planned in 1986 having as major objectives: (1) estimation of the prevalence of physical and mental impairments including Alzheimer's and other dementing diseases among persons in each of three age groups (55-64 yrs/65-74 years/ and 75 yrs. and older) in the Jing-An district of Shanghai, China, and (2) identification of approximately 1,400 persons in each of these 3 age groups for future determination of the incidence of these conditions. Jing-An, is one of twelve districts comprising the city of Shanghai, and was chosen as the target area because of its relatively large and stable population of elderly and its proximity to the Shanghai Institute of Mental Health which was responsible for the field work. Findings from this study have been discussed by Zhang *et al.* (1990) and by Yu *et al.* (1989). Methodological issues have been discussed by Levy *et al.* (1988 and 1989).

The clusters in this survey are administrative entities called *neighborhood groups* consisting of geographically contiguous households having a well defined social and political structure. The strategy was to involve the leaders of neighborhood groups selected in the sample in the identification and recruitment of eligible persons. At the time of the planning of the survey, there were 4,066 neighborhood groups within the Jing-An District. This particular population of aging and elderly Chinese generally had a low level of education and had experienced in their lifetimes repeated periods of political upheaval and repression (*e.g.*, the Warlords, the Japanese invasion, the Cultural Revolution), where being singled out or selected often had adverse consequences. For these reasons, it was felt strongly, especially by the local Chinese members of the research team who were most familiar with the target population, that any attempt to subsample persons in the target age groups within neighborhood groups that fall into the sample would compromise response rates and overall cooperation.

Restricted to single stage cluster sampling and faced with a very tight deadline for designing the sample, the member of the study team responsible for the sample design (PSL) proposed a heuristic method that would result with reasonable certainty in the identification of 1,400 individuals within each of the three target age groups. The resulting design was essentially a TSSCS in which 446 neighborhood groups were sampled. For details of this design, the reader is referred to the publications on the SSADD cited above. It should be emphasized that the resulting design was chosen purely on heuristic grounds and long before the theory behind this methodology was developed.

Of the 446 neighborhood groups sampled, 149 were designated as type 1, and 136 of these contained at least 1 person in the target age group (55 years and above). Since only the type 1 clusters have as eligible respondents all persons in each of the 3 target age groups, they can be used to estimate all of the parameters needed to evaluate the cost effectiveness of TSSCS relative to OSSCS. In the ensuing discussion, we will use the data from these 136 clusters to illustrate numerically how, on the basis of available “pilot” data, comparisons can be made between OSSCS and TSSCS with respect to cost effectiveness. From this sample of 136 clusters, we have for each domain, h , estimates of relevant parameters as shown below:

Age	\bar{N}_h	V_{N_h}	\bar{X}_h	V_{hx}	$\sum_{k < h} \sigma_{hkx}$
55-64	10.985	.485	.125	2.991	0.000
65-74	8.088	.513	.360	2.357	0.190
75 +	3.478	.643	.456	1.665	0.296.

If we wish to identify with 95% confidence at least 1,400 persons in each age group, then from relation (1) and the data shown above, we would have

$$A_1 = 1.645 \times 0.485/2 = 0.3989$$

and

$$m'_1 = \left[0.3989 + \left((0.3989)^2 + \frac{1,400}{10.985} \right)^{1/2} \right]^2 = 136.78 \approx 137.$$

Similarly, $m'_2 \approx 185$, and $m'_3 \approx 419$.

Let us suppose that for each of the three age groups, we wish to estimate with 80% confidence to within 30% of its true value the proportion, \bar{X}_h , of persons showing evidence of cognitive dysfunction as judged by a score below 18 on the Mini Mental State Examination (MMSE), which is a screening test for cognitive dysfunction. From these same data, we have the following estimates of the parameters necessary to determine the number of sample clusters required to meet this specification:

From relation (2), with $M = 4,066$, $\epsilon = 0.30$, and $z_{1-\alpha/2} = 1.28$, we have the following values of m''_h :

$$m''_1 = 157; \quad m''_2 = 99; \quad m''_3 = 50$$

and from relation (3), the number, m_h , of clusters required to satisfy both conditions in each domain is given by:

$$m_1 = \max(137, 157) = 157; \quad m_2 = \max(185, 99) = 185; \quad m_3 = \max(419, 50) = 419.$$

Thus, for an OSSCS design to satisfy both specifications, the number, m , of clusters required to be sampled would be 419. Likewise, a TSSCS design having 157 type 1, 28 type 2, and 234 type 3 sample clusters would satisfy both requirements.

The cost components, C_1 and C_2 , expressed in person hours, are estimated to be 20 and 2 respectively. The relatively high cost component, C_1 , associated with clusters is due to the fact that once a neighborhood group is selected in the sample, many hours must be spent obtaining the list of households and persons from a central bureau and enlisting the support of the neighborhood group leaders. The cost component, C_2 , of 2 person hours associated with individuals involves primarily interview and call-back activities. Thus, the field costs, C_0 , associated with an OSSCS design that satisfies both specifications is (from relation (7)) 27,278 person hours as compared to a cost of 17,737 person hours (from relation (8)) associated with a TSSCS design that satisfies both specifications. This represents a 35% savings in field costs, which is substantial.

From relation (9), we calculate that the estimate, x'_{ord} , of the number of persons over all 3 age groups having evidence of a cognitive disorder based on an OSSCS of 419 sample clusters would have variance equal to 70,844, whereas x'_{tel} , the analogous estimate based on a TSSCS also with 419 clusters, would have variance equal to 122,744, which is 42% greater than the variance of the OSSCS estimate. However, an OSSCS design having the same field costs as a TSSCS design based on 419 sample clusters would permit only 208 clusters to be sampled (relation (11)). The variance of x'_{ord} based on an OSSCS design with 208 sample clusters would be estimated to be 141,733, which is 15% higher than the variance of the analogous TSSCS estimate having the same field cost. Also, the OSSCS design having 208 sample clusters would not satisfy the two specifications placed on the estimates.

6. DISCUSSION

The methodology, TSSCS, discussed here and in earlier publications, arose from a situation in which cluster sampling was clearly indicated but a definite "red light" was given to any subsampling within clusters. For the Shanghai Survey of Alzheimer's Disease and Dementia considered here, the two major objectives were to identify a certain number of individuals within each of 3 domains (age groups in this instance) and to obtain domain specific estimates meeting certain specifications pertaining to precision. Based on results presented above for this particular survey, it appears that this method could result in considerable savings in field costs without compromising objectives.

One might raise questions concerning the general applicability of this methodology. It would be of use primarily in situations where it is either not feasible or too costly to subsample clusters and the individuals do not have to be screened to determine whether they belong to one of the target domains (in the SSADD, the leadership of the sample neighborhood groups provided a list of all persons in the neighborhood group along with information on data of birth). Such scenarios may occur, for example, in surveys where data are abstracted from records by personnel sufficiently familiar with the records to abstract information, but not considered capable of sampling the records without expensive supervision. Again, in such situations, TSSCS may provide a reasonable alternative.

ACKNOWLEDGEMENTS

Research on this report was supported, in part, by grant number 1 RO1 AG10327-01 from the National Institute on Aging.

REFERENCES

- KANNEL, W.B. (1966). An epidemiological study of cerebrovascular disease. In *Fifth Conference on Cerebrovascular Diseases*, (Eds. R.G. Sickert and J.P. Whisnart). New York: Grune and Stratton.
- LEVY, P.S., YU, E.S.H., LIU, W.T., ZHANG, M., WANG, Z., WONG, S., and KATZMAN, R. (1988). Variation on single stage cluster sampling used in a survey of elderly people in Shanghai. *International Journal of Epidemiology*, 17, 931-933.
- LEVY, P.S., YU, E.S.H., LIU, W.T., WONG, S., ZHANG, M., WANG, Z., and KATZMAN, R. (1989). Single-stage cluster sampling with a telescopic respondent rule: A variation motivated by a survey of dementia in elderly residents of Shanghai. *Statistics in Medicine*, 8, 1537-1544.
- LEVY, P.S., and LEMESHOW, S. (1991). *Sampling of Populations: Methods and Applications*. New York: John Wiley and Sons.
- YU, E.S.H., LIU, W.T., LEVY, P.S., ZHANG, M., KATZMAN, R., LUNG, C.T., WONG, S., and WANG, Z. (1989). Cognitive impairment among elderly in Shanghai, China. *Journal of Gerontology Social Sciences*, 44, 97-106.
- ZHANG, M., KATZMAN, R., SALMON, D., JIN, H., CAI, G., WANG, Z., QU, G., GRANT, I., YU, E.S.H., LEVY, P.S., KLAUBER, M.R., and LIU, W.T. (1990). The prevalence of dementia and Alzheimer's disease in Shanghai China: Impact of age, gender, and education. *Annals of Neurology*, 27, 428-437.

JOURNAL OF OFFICIAL STATISTICS

An International Review Published by Statistics Sweden

JOS is a scholarly quarterly that specializes in statistical methodology and applications. Survey methodology and other issues pertinent to the production of statistics at national offices and other statistical organizations are emphasized. All manuscripts are rigorously reviewed by independent referees and members of the Editorial Board.

Contents JOS 1991, Volume 7, Number 4

A Framework for Analyzing Categorical Survey Data with Nonresponse <i>David A. Binder</i>	393
Inference with Survey Weights <i>Roderick J.A. Little</i>	405
Cluster Analysis in International Comparisons <i>György Szilágyi</i>	425
A Multivariate Time Series Analysis of Fertility, Adult Mortality, and Real Wages in Sweden 1751-1850: A Comparison of Two Different Approaches <i>Mats Hagnell</i>	437
Census by Questionnaire - Census by Registers and Administrative Records: The Experience of Finland <i>Pekka Myrskylä</i>	457
A Comparison of the Missing-Data Treatments in the Post-Enumeration Program <i>Joseph L. Schafer</i>	475
Miscellanea	
Natural Resource and Environmental Accounting in the National Accounts <i>Lindy H. Ingham</i>	499
In Other Journals	515
Special Notes	519
Book Reviews	521
Index to Volume 7, 1991	531

All inquiries about submissions and subscriptions should be directed to the Chief Editor:

Lars Lyberg, U/SFI, Statistics Sweden, S-115 81 Stockholm, Sweden

Applied Statistics

JOURNAL OF THE ROYAL STATISTICAL SOCIETY (SERIES C)

CONTENTS

Volume 41, No. 2, 1992

	<i>Page</i>
Analysis of ordinal paired comparison data <i>A. Agresti</i>	287
Robust procedures for drug combination problems with quantal responses <i>T. J. Vidmar, J. W. McKean and T. P. Hettmansperger</i>	299
Pseudofactors: normal use to improve design and facilitate analysis <i>H. Monod and R. A. Bailey</i>	317
Adaptive rejection sampling for Gibbs sampling <i>W. R. Gilks and P. Wild</i>	337
A note on Wadley's problem with overdispersion <i>B. J. T. Morgan and D. M. Smith</i>	349
Sequential application of Wilks's multivariate outlier test <i>C. Caroni and P. Prescott</i>	355
Influence in correspondence analysis <i>P. Pack and I. T. Jolliffe</i>	365
A reduced rank regression model for local variation in solar radiation <i>C. A. Glasbey</i>	381
Hierarchical Bayesian analysis of changepoint problems <i>B. P. Carlin, A. E. Gelfand and A. F. M. Smith</i>	389
<i>General Interest Section</i>	
Interlaboratory comparisons: round robins with random effects <i>M. J. Crowder</i>	409
<i>Letter to the Editors</i>	427
<i>Book Reviews</i>	429
<i>Statistical Software Reviews</i>	
DYMO	436
SPSS/PC+	438
<i>Statistical Algorithms</i>	
AS 273 Comparing subsets of regressor variables <i>A. J. Miller</i>	443
AS 274 Least squares routines to supplement those of Gentleman <i>A. J. Miller</i>	458
AS 275 Computing the non-central χ^2 distribution function <i>C. G. Ding</i>	478
AS 276 Normal combinatoric classification <i>C. L. Dunn</i>	483
<i>Remark</i>	
AS R89 A remark on Algorithm AS 76: An integral useful in calculating non-central t and bivariate normal probabilities <i>P. W. Goedhart and M. J. W. Jansen</i>	496