Journal of Official Statistics
Vol. 3, No. 2, 1987, pp. 155–168

Interviewer-Respondent Bias Resulting from Adding Supplemental Questions

Gary M. Shapiro[1]

Abstract: Four cases are presented in which there were seemingly innocuous questionnaire changes, i.e., the addition of supplementary questions. In each case, there was a measurable impact on the basic results and evidence that respondent or interviewer bias was affected. In three cases, questions asked before the supplementary questions, were affected. The results and alternative explanations are discussed in detail, together with some implications for questionnaire design.

Key words: Questionnaire; conditioning effect; response bias; probing questions.

1. Introduction

This paper concerns four cases where supplementary questions have been added to an existing questionnaire. Although there were no advanced controlled experiments, there was a measurable impact on the survey results and evidence that respondent or interviewer bias was affected by the supplementary questions. In three of these cases, the addition of supplementary questions affected answers to questions asked earlier in the interview. Section 2 discusses the four cases in detail and offers alternative explanations for the results such as conditioning effects on respondents and various changes in interviewer behavior. Section 3 discusses the implications of the results for questionnaire design, namely that for one-time surveys, additional side questions may be needed to obtain good results, and that for continuing surveys, more care must be given to the addition of supplementary questions to avoid biased time series.

Related previous findings and experiments are now dicussed. It is well-known that question wording, ordering, etc. can affect survey responses. For example Kahn and Cannell (1957) discuss the effects of patterning and sequencing of questions. Most of the empirical work in the area of questionnaire design has been with "soft" attitudinal questions. Kalton et al. (1978) provides a good literature review as well as results from a new set of experiments on questionnaire design that show the substantial effects of questionnaire wording and context. Belson (1981) performed a detailed study to investigate respondent misunderstanding of 29 questions about television watching; many of these questions were

[1] Assistant Division Chief for Survey Design, Statistical Methods Division, Bureau of the Census, Washington, DC 20233, U.S.A.

Acknowledgements: This paper is an adaptation of an earlier paper (Gibson et al. (1978)). Much of the credit for this paper belongs to my co-authors of the earlier paper: Christina O. Gibson, Linda R. Murphy, and Gary J. Stanko. Any errors, however, are strictly my responsibility. I would also like to thank the two referees who provided comments that significantly improved this paper. Finally, many thanks to Edith Oechsler for the typing of the paper.

not attitudinal. He found that question wording contributed to substantial misunderstandings. Only 29 % of respondents interpreted a question within permissible limits across all questions (Belson (1981, p. 350)). Belson (1966) showed that response distributions for attitude scaling can be affected by an interviewer's reading the categories from positive to negative or from negative to positive. For earnings data, Borus (1970) showed that estimates were affected by whether general questions were asked or whether a longer work history approach was taken. McFarland (1981) showed that answers to attitude questions are affected by whether specific questions precede or follow general questions. Laurent (1972) found that the length of questions can also have an affect. Some careful experimentation by Presser and Schuman (1975) showed how question wording can affect attitude responses.

Several authors have considered the effect of the context and positioning of questions and it is these issues that are particularly germane to the results of this paper. Turner and Krauss (1978) speculated that different estimates of levels in public confidence in national institutions resulted from whether the relevant questions were placed at the beginning of the questionnaire or not. Metzner and Mann (1953) found no evidence of effects from grouping a set of questions together rather than interspersing the questions throughout the questionnaire. Bradburn and Mason (1964) also found that changing the order of questionnaire sections did not appear to have an effect. Sudman and Bradburn (1973), however, found that estimates produced from consumer expenditure questions were affected by placing the questions at the beginning, middle, or end of the interview. For questions requiring recall of 14 weeks or more, response was lower when the questions occurred towards the end of the interview. However, for recall of less than two weeks, question placement had no apparent affect. Herzog and Bachman (1981) found that the length of a questionnaire given to students in a classroom did not generally affect results. There was, however, one exception; for a longer questionnaire, a related set of questions towards the end of a questionnaire are more likely to be all answered identically. Noelle-Neumann (1976) found results similar to those of this paper. "She (Noelle-Neumann) provides some data indicating that following up a particular response with another question, e.g., "If yes – In what way?,' may reduce the number of people who will say "yes' or give any opinion at all. It is not clear whether this effect is provided by the respondents' desire to avoid the follow-up question (and thus reduce the burden of answering) or by the interviewer's cluing respondents that answering a particular way will prolong the interview." (Bradburn (1978).)

The four cases discussed in this paper are consistent with the findings of Turner and Krauss, of Sudman and Bradburn, and particularly of Noelle-Neumann. The first two sets of authors found that question placement affected results. Noelle-Neumann and this paper address the issue of whether the inclusion or exclusion of questions affected the answers to other questions. All four cases in this paper deal with relatively "hard" data (criminal victimizations, unemployment, not in the labor force, and acute health conditions). Particularly important is that in three of the cases the addition of supplementary questions affected answers to questions asked *earlier* in the interview. The effect on the data from each of the four sets of supplementary questions is somewhat surprising.

2. The Four Cases

Case A. Effect of an Attitude Supplement on National Crime Survey (NCS)-Cities Sample Victimization Data

In each of the cities in the National Crime Survey (NCS), a lengthy series of attitude questions was asked of respondents in a random half of sample units as a supplement to the basic crime victimization questions. The NCS-Cities was conducted by the U.S. Bureau of the Census for the Law Enforcement Assistance Administration, the predecessor organization for the Bureau of Justice Statistics. Information was collected from victims who were identified by screening a general population sample in selected large central cities. Respondents were interviewed about experiences in the preceding 12 months with selected crimes of violence and theft, including events not reported to police as well as reported crimes. See U.S. Department of Justice (1976a and 1976b) for further information. For each respondent, the added attitude questions were purposely asked before the victimization questions to prevent influencing or conditioning the attitude responses, since the attitude questionnaire included such topics as neighborhood safety, behavior patterns, and opinions regarding local police, crime trends, and news coverage of crime.

The attitude supplement lengthened the total interview time considerably for the half-sample chosen for the supplement. In fact, concern arose that the additional time might have increased respondent fatigue, which in turn might have resulted in fewer reports of victimization. To determine whether this hypothesized effect actually occurred, special tabulations were produced for each of the 13 cities surveyed in 1975, showing personal and property crime victimization rates separately for each of the half-samples. Personal crimes included: completed and attempted assault, rape, robbery, purse snatching, and pocket picking. Property crimes included completed and attempted burglary, larceny (both household and personal), and auto theft.

The data in Table 1 on the following page show relative differences in victimization rates reported in the half-samples containing the attitude supplement versus those without the supplement. These data clearly indicate that the inclusion of the attitude supplement in the NCS interview had a statistically significant impact on the victimization rates obtained. Additionally, the differences are not in the hypothesized direction of lower reporting of crime incidents. Rather, in almost every city, the victimization rate measured for the half-sample in which the attitude questions were asked was higher than for the other half-sample in which they were not asked, both for property and personal crimes.[2]

The most plausible explanation for these results is that asking attitude questions before the victimization questions had a conditioning effect, whereby the respondents' awareness or memory regarding victimization experience was stimulated. Motivational factors may also play a role here. Kahn and Cannell (1957) point out that a desire to have an influence can be an impetus for respondent motivation. In this case, asking attitude questions may heighten some respondents' feelings that their answers may influence government policies and actions. Increased motivation and desire to be accommodating is likely to lead to fuller reporting of actual victimizations, or to exaggerating or fabricating victimizations. A third explanation, which combines elements of both of the above, is that administering the attitude questions serves to stimulate both memory and motivation. This results in the reporting of more real (not fabricated) incidents that actually happened before the reference period, i.e., increased telescoping. It is important to note that if the second or third explanation is true, the higher victimization

[2] A similar comparison was made between two half-samples for an additional 13 NCS cities surveyed in 1974 with similar results. A more detailed examination of differences is contained in Cowan, et al. (1978).

Table 1. Victimization Rates and Relative Differences[1] in Rates for the Half-Samples With and Without the Attitude Supplement

City	Rate for half-sample with the attitude supplement				Rate for half-sample without the attitude supplement				Relative differences in rates between the half-samples			
	Personal crimes		Property crimes		Personal crimes		Property crimes		Personal crimes		Property crimes	
Chicago	8.40	(0.27)	42.95	(0.76)	7.09	(0.25)	38.97	(0.75)	+18.47**	(5.61)	+10.21**	(2.88)
Philadelphia	6.64	(0.26)	37.83	(0.73)	5.63	(0.23)	35.72	(0.72)	+17.99**	(6.67)	+ 5.93**	(2.95)
New York	6.16	(0.24)	27.43	(0.69)	5.36	(0.22)	23.89	(0.66)	+14.97**	(6.59)	+14.82**	(4.27)
Los Angeles	7.50	(0.28)	60.43	(0.76)	5.81	(0.25)	51.22	(0.77)	+29.02**	(7.37)	+17.97**	(2.32)
Detroit	8.93	(0.30)	52.95	(0.80)	8.21	(0.29)	48.66	(0.80)	+ 8.82*	(5.29)	+ 8.82**	(2.42)
Portland, Oregon	8.40	(0.30)	69.24	(0.71)	6.77	(0.27)	63.70	(0.74)	+24.09**	(6.55)	+ 8.68**	(1.68)
Denver	8.67	(0.29)	68.98	(0.72)	6.76	(0.26)	59.99	(0.75)	+28.12**	(6.49)	+14.97**	(1.87)
Baltimore	10.67	(0.31)	49.91	(0.75)	8.67	(0.28)	45.60	(0.74)	+23.03**	(5.26)	+ 9.45**	(2.42)
Atlanta	6.06	(0.27)	49.81	(0.79)	4.54	(0.24)	44.35	(0.78)	+33.39**	(9.17)	+12.31**	(2.66)
Dallas	6.32	(0.23)	63.91	(0.74)	4.62	(0.20)	54.38	(0.76)	+36.84**	(7.78)	+17.53**	(2.13)
Cleveland	8.34	(0.29)	50.90	(0.74)	6.90	(0.27)	44.39	(0.75)	+20.90**	(6.30)	+14.67**	(2.56)
St. Louis	5.95	(0.25)	49.45	(0.79)	5.51	(0.24)	40.02	(0.77)	+ 7.85	(6.51)	+23.57**	(3.10)
Newark	5.30	(0.22)	27.20	(0.67)	4.38	(0.20)	24.96	(0.64)	+21.02**	(7.38)	+ 8.99**	(3.87)

[1] Computed as: $\left(\frac{\text{Victimization rate for half-sample with the attitude supplement}}{\text{Victimization rate for half-sample without the attitude supplement}} - 1\right) \times 100$

( ) = One standard error.
* Indicates statistically significant difference at the 10 percent level between the rates for each of the half-samples.
** Indicates statistically significant difference at the 5 percent level between the rates for each of the half-samples.

Source: U.S. Department of Justice (1976a and 1976b).

Table 2. Incidence[1] of Acute Conditions and Number of Acute Conditions per 100 Persons per Year by Sex and Age

All acute conditions	1971		1972		1973[2]		1974[2]		1975	
Incidence in thousands										
Both sexes	442 203	(7 998)	448 607	(8 093)	360 448	(6 784)	364 278	(6 841)	443 119	(8 012)
Male	202 787	(4 387)	202 558	(4 383)	170 046	(3 871)	171 661	(3 896)	204 920	(4 420)
Female	239 416	(4 954)	246 049	(5 056)	190 402	(4 193)	192 617	(4 228)	238 199	(4 936)
Number of acute conditions per 100 persons per year										
Both sexes	218.5	(4.0)	219.7	(4.0)	175.1	(3.3)	175.7	(3.3)	212,0	(3.8)
Male	207.8	(4.5)	205.8	(4.5)	171.3	(3.9)	171.6	(3.9)	203.2	(4.4)
Female	228.5	(4.7)	232.8	(4.8)	178.7	(3.9)	179.5	(3.9)	220.1	(4.6)
Incidence in thousands										
All ages	442 203	(7 998)	448 607	(8 093)	360 448	(6 784)	364 278	(6 841)	443 119	(8 012)
Under 6 years	79 707	(2 351)	76 696	(2 295)	62 089	(2 019)	61 121	(2 000)	75 731	(2 278)
6–16 years	133 582	(3 280)	120 989	(3 070)	100 432	(2 719)	102 172	(2 750)	112 612	(2 929)
17–44 years	154 029	(3 614)	166 144	(3 808)	136 560	(3 329)	141 483	(3 410)	176 558	(3 974)
45 years and over	74 885	(2 262)	84 778	(2 443)	61 368	(2 005)	59 503	(1 969)	78 218	(2 323)
Number of acute conditions per 100 persons per year										
All ages	218.5	(4.0)	219.7	(4.0)	175.1	(3.3)	175.7	(3.3)	212.0	(3.8)
Under 6 years	372.7	(11.0)	369.9	(11.1)	304.5	(9.9)	309.0	(10.1)	388.1	(11.7)
6–16 years	295.8	(7.3)	274.2	(7.0)	230.3	(6.2)	236.7	(6.4)	265.4	(6.9)
17–44 years	206.2	(4.8)	215.4	(4.9)	172.8	(4.2)	175.1	(4.2)	213.4	(4.8)
45 years and over	122.5	(3.7)	136.4	(3.9)	97.7	(3.2)	93.6	(3.1)	121.5	(3.6)
Percent of all reported acute conditions reported as beginning in the two-week reference period[3]	75.4		73.6		56.6		57.9		67.6	

[1] Estimates of the incidence of acute conditions include, with certain exceptions (i.e., conditions that are always classified as chronic), only those conditions which had started within the two weeks prior to the week of interview and which had involved either medical attention or restricted activity.
[2] The Acute Conditions supplement was asked this year.
[3] Based on unweighted numbers.

( ) = One standard error.

Source: U.S. Department of Health, Education, and Welfare (1973–1977).

rates may not indicate better data but simply an undesirable response bias. On the other hand, the first explanation points to a response bias in the half-sample not asked the attitude questions.

Case B. Impact of the Acute Conditions Supplement on the Number of Acute Conditions Reported in the Health Interview Survey (HIS)

Special supplementary questions about acute conditions that were first noticed within the two weeks prior to the week of interview was included in the Health Interview Survey (HIS) in 1973 and 1974. The HIS is conducted by the U.S. Bureau of Census for the National Center for Health Statistics. Information is collected on personal and demographic characteristics, illnesses, injuries, impairments, chronic and acute conditions, and other health topics from weekly samples of households, from which the data are additive over time. An acute condition is defined as a condition which has lasted less than three months and which has involved either medical attention or restricted activity. See Kovar and Poe (1985) for further information. For the Acute Condition Supplement, the interviewer was usually required to personally interview the person who had the acute condition and ask at least 55 additional questions.

The effect of including the supplemental questions on acute conditions was dramatic and is shown in Table 2 on page 159. (U.S. Department of Health, Education, and Welfare (1973–1977).) Incidence and number of acute conditions were substantially lower for 1973 and 1974 (the two years the supplement was asked) than for 1971, 1972, and 1975. The number of acute conditions per 100 persons declined by 20.3 percent from 1972 to 1973 and increased by 20.7 percent from 1974 to 1975. Other acute conditions lasting less than three months are reported in the early part of the main questionnaire. Interestingly, the relative number of such conditions did not change from 1971 through 1975; it remained at about 120 acute conditions per 100 sample persons per year.

Even though these results were not obtained under controlled experimental conditions, the differences are so striking that it seems highly unlikely that real changes in the nation's health account for these changes. The only change in the questionnaire relating to acute conditions for 1973 and 1974 was the requirement to complete an additional supplement about conditions that were first noticed during the two-week reference period. The effect of this change had been expected to be minimal. Additional analysis showed that the drop from 1972 to 1973 could not be associated with any particular segment of the population, and it did not appear that there was any appreciable reduction in medical attention or activity restriction that would account for the lower incidence rates. The incidence rates of 175.1 and 175.7 per 100 persons per year for 1973 and 1974, respectively, were the lowest reported in the HIS since its inception in July 1957.

There are several possible explanations for these results. Noelle-Neumann (1976), who found similar effects from the addition of a simple follow-up question, suggested two possible explanations (mentioned in Section 1 of this paper). There may in fact be two or more reasons for the acute condition results. One such reason is that more care was exercised during the two years the supplement was administered in finding out if an acute condition actually did start within the two-week reference period as opposed to starting prior to the reference period. It is difficult for respondents to remember exactly when an acute condition started. Therefore, because of the additional requirements for completing a supplement, more extensive probing of respondents with greater use of memory aids could have resulted in more precise pinpointing of the date of onset of the acute condition. However, as mentioned earlier, other acute conditions lasting less than three months did not increase during this time period. A more plausible explanation is that in some cases the interviewer did not want to burden him/herself, or to burden the respondent with more questions and therefore classified some respondents incorrectly.

A second plausible reason for the results is that the presence of the long supplement caused interviewers to rush through the questions to complete the interview more quickly. This in turn might have made respondents feel that the interviewers wanted the answers to be quick rather than well-prepared. If interviewers had been carefully questioned on their experiences in 1973 or 1974, this might have helped in determining what the real cause or causes were. Unfortunately, this was not done, neither in this case nor in the other cases discussed in this paper.

Case C. Impact of the Not-in-the-Labor Force Supplement on Labor Force Data

The Survey of Work History and Job Search Activities of Persons Not in the Labor Force was a Current Population Survey (CPS) supplement administered to a portion of the two outgoing rotation groups of the eight rotation groups in both July and August 1976. The Current Population Survey is conducted by the U.S. Bureau of the Census for the Bureau of Labor Statistics. The survey is conducted monthly to provide estimates of employment, unemployment, and other labor force characteristics. Each CPS sample is divided into eight different parts called rotation groups. Each rotation group is a scientifically selected subsample of the complete sample, and each month one of the rotation groups is introduced into the survey. The households in a rotation group are interviewed for four consecutive months, dropped from the survey for eight months, and then interviewed again for four more consecutive months. The month-in-sample designation indicates how many months a rotation group has been in the survey. One phenomenon associated with this rotation pattern is "rotation group bias" where the level of estimates for many characteristics varies depending on the number of months the household has been in sample. For example, the unemployment rate is usually highest for first month-in-sample households, and then continually drops off for later interviews, except for some rebounding in the fourth, fifth, and eighth months in sample. For further information, see Hanson (1978).

The supplement was completed immediately after the basic CPS interview when possible; otherwise, a self-enumeration questionnaire was left or mailed for the respondent to return by mail. Up to 60 additional questions, depending on the response pattern, had to be answered. The additional questions took an average of 20 minutes to complete. The universe eligible for the supplement consisted of persons 14 years old and over who were not in the labor force, but who wanted work. Since this group can be identified every month from the basic CPS questionnaire, it was relatively easy to determine that the universe interviewed for the supplement was about half of what we had expected it to be. The noninterview rate on the supplement was high, about 20 percent, but this cannot explain that so few sample persons were initially identified by the interviewers as eligible for the universe. The examination of two separate sets of data led us to this conclusion.

First we looked at the number of persons reported as not in the labor force and intending to look for work in the next 12 months and found that it decreased significantly in July and August 1976 from the June 1976 count (about 27 percent) and from the July and August 1975 figures (about 25 percent). See Table 3 (Taylor (1976)). The seasonally adjusted number of persons who "want a job now" in 1976 dropped 20.0 percent from the second to the third quarter and increased 25.3

Table 3. Persons Aged 16 Years and Over Not in the Labor Force Who Intend to Look for Work in the Next 12 Months for Certain Months of Certain Years (in thousands)

Month	Total persons not in the labor force who intend to look for work in the next 12 months											
	1975		Month-to-month percent change		1976		Month-to-month percent change		1977		Month-to-month percent change	
April	9 701	(124)			9 637	(125)			10 004	(11)		
			−12.5**	(0.02)			−14.0**	(0.02)			− 8.0**	(0.02)
May	8 488	(140)			8 291	(141)			9 313	(130)		
			−12.9**	(0.02)			− 8.7**	(0.02)			−15.6**	(0.02)
June	7 397	(147)			7 571	(146)			7 856	(145)		
			− 3.3	(0.03)			−28.0**	(0.02)[1]			+ 0.5	(0.03)
July	7 151	(148)			5 452	(149)			7 897	(144)		
			+ 6.5**	(0.03)			+ 3.1	(0.04)			+ 2.8	(0.03)
August	7 616	(146)			5 620	(149)			8 122	(143)		
			+27.9**	(0.03)			+76.4**	(0.05)			+26.5**	(0.03)
September	9 739	(124)			9 912	(121)			10 278	(114)		

[1] Represents figures for the two months in which the supplement was administered (see text).
( ) = One standard error.
* Indicates statistically significant difference at the 10 percent level.
** Indicates statistically significant difference at the 5 percent level.

Source: Taylor (1976).

Table 4. Persons Aged 16 Years and Over Not in the Labor Force Who Want a Regular Job Now, Either Full- or Part-Time Quarterly Averages, Seasonally Adjusted (in thousands)

Quarter	1975		Quarter-to-quarter percent change		1976		Quarter-to-quarter percent change		1977		Quarter-to-quarter percent change	
January–March	5 211	(91)			5 388	(92)			5 663	(93)		
			−2.4	(0.02)			+ 0.7	(0.02)			+1.7	(0.02)
April–June	5 084	(91)			5 426	(92)			5 762	(93)		
			+5.3**	(0.03)			−20.0**	(0.02) [1]			+2.6	(0.02)
July–September	5 354	(92)			4 339	(87)			5 909	(94)		
			−1.8	(0.02)			+25.3**	(0.03)			−5.8**	(0.02)
October–December	5 256	(91)			5 436	(92)			5 565	(93)		
			+2.5	(0.02)			+4.2*	(0.02)				

[1] The supplement was administered in two of the three months in this quarter.
( ) = One standard error.
* Indicates statistically significant difference at the 10 percent level.
** Indicates statistically significant difference at the 5 percent level.

Source: U.S. Department of Labor (1978).

Table 5. Eligibility for Not-in-the Labor Force Supplement in the Original Interview and the Reinterview

	July and August 1976	Reinterview
Not in labor force (NILF)	22 942	843
Sample count of respondents eligible for supplement	3 387	167
Eligible respondents as percentage of NILF	14.8	19.8

Source: Internal Bureau of the Census Data.

percent in the fourth quarter; this pattern was not repeated in 1975 or 1977 (see Table 4) (U.S. Department of Labor (1978)).

We then looked at data from the CPS Content Reinterview and Reconciliation Program[3]. This indicate that too few persons were identified in the original interview as eligible for the supplement. For the original interview in July and August 14.8 percent of those persons not in the labor force were eligible for the supplement. For the reinterview, 19.8 percent were eligible. (See Table 5.)

Thus, there was overwhelming evidence that the existence of the supplement reduced the number of persons reported as not in the labor force and intending to look for work. The interviewers could have used more probes to determine whether the respondent truly wanted a job now or in the next 12 months. Maybe the interviewer did not wish to ask more burdensome questions and changed the classification of the respondent accordingly. Another possibility is that the additional information obtained in the supplement resulted in interviewers reclassifying some respondents by "correcting" original answers, an action they were instructed not to make.

[3] A portion of the sample is re-asked the basic labor force questions the week after interview week by senior interviewers for quality control purposes. See Bailey et al. (1968).

Case D. Impact of the Discouraged Workers Question on CPS Unemployment Data

The CPS interviewer asks a series of questions (known as the discouraged workers questions) only to those people who have answered in previous questions that they are neither working, nor looking for work, and therefore are not in the labor force. Furthermore, questions are asked only of people in two rotation groups out of eight in the CPS. This series was not asked in the CPS prior to January 1967. From January 1967 through December 1969, the series was asked in respondents' first and fifth interviews. From January 1970 to the present, the series has been asked the fourth and eighth interviews only. Table 6 shows the impact of the discouraged workers questions on labor force status estimates for individual months-in-sample (Bailar (1975)). The rotation group bias indices in the table were computed for the i^{th} month-in-sample as follows:

$$\frac{\text{Estimate for the } i^{\text{th}} \text{ month-in-sample}}{\text{Total estimate for time period / 8}} \times 100.$$

Figures in italics in the table are for those months-in-sample in which the discouraged workers questions were asked. The effect is largest for estimates of unemployment. This

Table 6. Rotation Group Indices in the CPS for Two Periods, 1968–69 (T1) and 1970–72 (T2), for Selected Characteristics

		Month in sample								Monthly average class size (000)	Standard error of index
		1	2	3	4	5	6	7	8		
Total population 16 and over											
Civilian labor force	T1	_102.3*_	100.3	99.8	_99.5_	_100.8*_	99.3	99.1	_99.0_	80 340	0.3
	T2	101.6	100.0	99.6	_100.3*_	100.0	99.1	99.2	_100.0**_	84 654	0.2
Employed	T1	_101.6*_	100.2	99.9	_99.8_	_100.4*_	99.4	99.4	_99.3_	77 285	0.3
	T2	101.1	100.0	99.7	_100.3*_	99.9	99.4	99.5	_100.1*_	79 913	0.2
Unemployed	T1	_120.0**_	101.5	96.4	_92.8_	_109.3*_	96.5	92.6	_91.0_	3 055	2.4
	T2	109.2	100.3	98.1	_101.2**_	102.3	96.7	94.1	_98.2**_	4 741	1.2
Males 16 and over											
Employed	T1	_100.9_	100.0	100.0	_99.8_	_100.2_	99.8	99.7	_99.7_	48 589	0.3
	T2	100.7	99.9	99.8	_100.2_	99.9	99.7	99.7	_100.2*_	49 637	0.2
Unemployed	T1	_114.1*_	102.6	98.0	_95.6_	_106.0*_	97.7	93.4	_92.6_	1 490	3.5
	T2	105.4	101.4	99.9	_101.6*_	100.3	98.0	95.6	_97.9*_	2 578	1.6
Females 16 and over											
Civilian labor force	T1	_104.0*_	100.6	99.5	_99.2_	_101.4*_	98.7	98.4	_98.1_	30 261	0.6
	T2	102.7	100.0	99.4	_100.4*_	100.2	98.6	98.6	_100.0*_	32 439	0.5
Unemployed	T1	_125.5**_	100.2	94.6	_90.5_	_112.5*_	95.1	91.8	_89.5_	1 564	2.8
	T2	113.8	99.0	95.9	_100.7**_	104.4	95.2	92.2	_98.6**_	2 163	1.5

* Indicates statistically significant difference at the 10 percent level between the index for T1 and T2.
** Indicates statistically significant difference at the 5 percent level between the index for T1 and T2.
Figures in italics are for those months-in-sample in which the discouraged workers questions were asked.

Source: Bailar (1975).

indicates that when the discouraged workers questions are asked, there is a substantial increase in the number of unemployed. The difference between 1968–1969 and 1970–1972 shown in the tables may reflect real differences over time in rotation group bias as well as the effect of the discouraged workers questions.

One explanation for these results is that interviewers appear to learn new information when asking these probing questions. As a result, they go back and reclassify people even though they are explicitly instructed not to do so. Another explanation is that the household respondent becomes more accustomed answering questions about the labor force participation of subsequent household members after answering discouraged workers questions for one household member. Thus the household respondent answers differently after having already gone through the discouraged workers questions. It may also be, as suggested by Noelle-Neumann (1976), that interviewers handle the interview differently when they know that they will be asking the discouraged workers questions, and that this affects the classification of people as unemployed.

3. Implications

The above findings have two main applications. The first is when a questionnaire is initially designed for a survey, and the second is when a basic questionnaire is used in a recurring survey and supplementary questions are under consideration. Recall that this paper has discussed four cases in which the addition of supplementary questions has affected the answers to other questions. In the National Crime Survey, asking attitude questions relating to crime increased the reported criminal victimizations reported later on in the interview. In the Health Interview Survey, a set of supplementary questions for reported acute conditions decreased the number of acute conditions reported. In the Current Population Survey (CPS), a supplement in which questions on work history and job search activities, asked only of people who want work but are not in the labor force, resulted in fewer people being classified as wanting work but not in the labor force. Also, in the CPS, a set of questions is sometimes asked to those persons categorized as discouraged workers. When these questions are asked, the estimate of unemployed persons is higher.

When a questionnaire is initially developed and designed, a designer cannot simply derive what appears to be a straightforward set of questions and assume that they will yield reasonably valid answers. For example, if one is interested in accurate estimates of the unemployment rate, it may not suffice to ask a set of questions about recent activities. It may be necessary to structure a whole set of probing questions on side issues, where the answers to the probes have an impact on the classification of labor force status. (It may be, however, that certain probing questions will produce less accurate estimates.) As a second example, one may be interested in detailed information about housing and the attitudes of persons in certain types of housing, e.g., low rent or low value. Ordinarily, one would then design a questionnaire with a set of attitudinal questions asked only when a certain set of answers to the "basic" housing questions is obtained. One would assume this to yield good housing data and good attitudinal data. The HIS and CPS examples in this paper, however, would indicate that this is not the way to proceed. (One alternative might be to ask attitudinal questions of everyone.)

When adding a set of supplementary questions to a recurring survey, the problem is to retain a consistent time series so that estimates of change over time are unbiased. If there is less interest in unbiased estimates of level than of change, it might be undesirable to add a set of supplementary questions, even if it is known that it will improve the accuracy of the answers to the basic questions. There are several possible solutions. Additional questions could be administered to only a portion of the sample, small enough not to have an appreciable effect on the total data, or to some portion of the sample which may not be used to produce basic estimates. Or a different series of questions could be administered during each enumeration period. The effect would be to consistently produce positive results, if the assumption that the interaction produces better data is valid. If it is thought that interviewers might adjust for inconsistencies contrary to instructions (as suggested by the discouraged workers series of questions in the CPS), then contradictory responses should be allowed for in the wording of the questionnaire. To prevent interviewers from changing or incorrectly marking the answers to certain questions on the original questionnaire, and in doing so influencing the expected universe for the additional questions and causing bumps in the data series (as may have happened with the HIS and CPS Not-in-the-Labor-Force supplements), control over the distribution of the supplements must be removed from the interviewers' hands. Either all respondents must be asked supplemental questions so that there is no reason for interviewers to modify the answers, or the determination of the universe for the questions must be made independently of the interviewer, and possibly even without the interviewer's knowledge.

Ultimately, better data is needed (obtained through imbedded experiments) to give more precise estimates of the contributions of interviewers and respondents to the outcome of the data or of the interaction of the questions. This is especially true since the nature, magnitude, and direction of such nonsampling error cannot always be anticipated or predicted. We could then evaluate the trade-offs between damage to the basic survey data, the value of the additional information, the costs of other ways of collecting the data (e.g., a separate survey), and the procedures to lessen observed biases in a more scientific manner.

In conclusion, the findings presented here are rather disturbing. One knows that questions, questionnaire design, and interviewers can make a difference. However, on the surface, these cases entailed a set of relatively innocuous questionnaire changes; and yet when examined closely, quantitatively large nonsampling errors be came apparent. This suggests that there are other substantial nonsampling errors still hidden from view in these and other surveys, and superficially simple differences or changes in questionnaires or interviewing procedures can result in large biases.

4. References

Bailar, B. A. (1975): The Effects of Rotation Group Bias on Estimates from Panel Surveys. Journal of the American Statistical Association, 349, pp. 23–30.

Bailey, R. et al. (1968): The Current Population Survey Reinterview Program, January 1961 through December 1966. Bureau of the Census Technical Paper 19, Government Printing Office, Washington, DC.

Belson, W. A. (1981): The Design and Understanding of Survey Questions. Gower Publishing Co. Ltd, Aldershot, Hauts., England.

Belson, W. A. (1966): The Effects of Reversing the Presentation Order of Verbal Rating Scales. Journal of Advertising Research, Vol. 6, No. 4, pp. 1–11.

Borus, M. E. (1970): Response Error and Questioning Techniques in Surveys of Earnings Information. Journal of the American Statistical Association, Vol. 65, No. 330, pp. 566–575.

Bradburn, N. M. (1978): Respondent Burden. American Statistical Association. Proceedings of the Section on Survey Research Methods, pp. 35–40.

Bradburn, N. R. and Mason, W. M. (1964): The Effect of Question Order on Responses. Journal of Marketing Research, Vol. 1, No. 4, pp. 57–61.

Cowan, C. D., Murphy, L. R., and Wiener, J. (1978): Effects of Supplemental Questions on Victimization Estimates from the National Crime Survey. American Statistical Association. Proceedings of the Section on Survey Research Methods, pp. 277–282.

Gibson, C. O., Shapiro, G. M., Murphy, L. R., and Stanko, G. J. (1978): Interaction of Survey Questions as it Relates to Interviewer-Respondent Bias. American Statistical Association. Proceedings of the Section on Survey Research Methods, pp. 251–256.

Hanson, R. (1978): The Current Population Survey: Design and Methodology. Bureau of the Census, Technical Paper 40, Government Printing Office, Washington, DC.

Herzog, R. A., and Bachman, J. G. (1981): Effects of Questionnaire Length on Response Quality. Public Opinion Quarterly, 45, pp. 549–559.

Kahn, R. L. and Cannell, C. F. (1957): The Dynamics of Interviewing-Theory, Technique, and Cases. Chapters 2 and 6. John Wiley & Sons, New York, pp. 131–165.

Kalton, G., Collins, M., and Brook, L. (1978): Experiments in Wording Opinion Questions. Applied Statistics, 27, No. 2, pp. 149–161.

Kovar, M. G., and Poe, G. (1985): NHIS: Design 1973–1984 and Procedures 1975–1983. U.S. Department of Health and Human Services, National Center for Health Statistics, Series 1, No. 18, 1985.

Laurent, A. (1972): Effects of Question Length on Reporting Behavior in the Survey Interview. Journal of the American Statistical Association, Vol. 67, No. 338, pp. 298–305.

McFarland, S. G. (1981): Effects of Question Order on Survey Responses. Public Opinion Quarterly, 45, pp. 208–215.

Metzner, H. and Mann, F.C. (1953): Effects of Grouping Related Questions in Questionnaires. Public Opinion Quarterly, Vol. 17, No. 1, pp. 136–141.

Noelle-Neumann, E. (1976): Die Empfindlichkeit demoskopischer Messinstrumente. In Allesbacher Jahrbuch der Demoskopie. Verlag Fritz Molden, Wien.

Presser, S. and Schuman, H. (1975): Question Wording as an Independent Variable in Survey Analysis: A First Report. American Statistical Association, Proceedings of the Social Statistics Section, pp. 16–25.

Sudman, S. and Bradburn, N.M. (1973): Effects of Time and Memory Factors on Response in Surveys. Journal of the American Statistical Association, Vol. 68, No. 344, pp. 805–815.

Taylor, R. W. (1976): Status-Report – July/August 1976 Survey of Work History and Job Search Activities of Persons Not in the Labor Force. Internal Census Bureau Memorandum to B. Gregory Russell, December 10, 1976.

Turner, C. F. and Krauss, E. (1978): Fallible Indicators of the Subjective State of the Nation. American Psychologist, Vol. 33, No. 5, pp. 456–470.

U.S. Department of Health, Education, and Welfare (1973–1977): National Center of Health Statistics Series 10, Nos. 79, 85, 95, 100, 102, 115. Government Printing Office, Washington, DC.

U.S. Department of Justice (1976a): Law Enforcement Assistance Administration. Criminal Victimization Surveys in Eight American Cities. No. SD-NCS-C-5. Government Printing Office, Washington, DC.

U.S. Department of Justice (1976b): Law Enforcement Assistance Administration. Criminal Victimization Surveys in Chicago, Detroit, Los Angeles, New York, and Philadelphia. No. SD-NCS-C-6. Government Printing Office, Washington, DC.

U.S. Department of Labor, Bureau of Labor Statistics (1978): Employment and Earnings. Vol. 24, No. 1. Government Printing Office, Washington, DC.

Received May 1985
Revised May 1987

Journal of Official Statistics
Vol. 3, No. 2, 1987, pp. 169–181

Classification and Properties of Rotation Sampling Designs

Ronaldo Iachan[1] and Byron Jones[2]

Abstract: For repeated sampling in time, a general class of rotation sampling designs is introduced. The construction and classification of such designs are followed by a detailed examination of their properties. For this purpose, an additive model is proposed that takes into account period and sampling unit effects and permits efficiency comparisons.

Key words: Rotation sampling design; efficiency; variance of period changes; construction algorithm; classification; properties.

1. Introduction

In this paper we consider the construction, classification, and properties of a class of rotation sampling designs, drawing an analogy with incomplete block designs. Efficiency comparisons are made using a linear additive model for survey responses.

Rotation sampling designs perform well when one wishes to minimize sampling and nonsampling errors. As for sampling errors, current levels are best estimated (in the sense of smallest sampling variance) when there is some overlap between successive samples. Partially overlapping samples provide a useful compromise between the complete overlap (same unit interviewed over and over), required for the most precise estimates of period-to-period changes, and the drawing of a new sample (no overlap) needed for the most precise overall (aggregate) estimates. See Cochran (1977, pp. 344–347) for related arguments and some classical developments for sampling on two or more occasions with simple random sampling. With more than two occasions, finding optimal designs becomes extremely complicated (even in a restricted and approximate sense). Patterson (1950) and Eckler (1955) have developed the basic theory for this case. Extensions to more general designs and estimators are given in Ghangurde and Rao (1969), Jain (1981), and Wolter (1979). As for non-sampling errors, the rotation aspect of the design avoids excessive respondent burden and sample attrition that lead to increased response errors and nonresponse errors.

The model-based approach of this paper permits straightforward efficiency computations when more than two occasions are considered. Recommendations as to which

[1] Research Triangle Institute, Center for Statistics Research, Research Triangle Park, North Carolina 27709-2194, U.S.A.
[2] Mathematical Institute, The University, Canterbury, Kent, CT2 7NF, England.
Acknowledgements: We are most grateful to the Editor and the referees for their helpful suggestions which have improved the presentation of the paper.

design to use in what circumstances follow from this methodology, so that design choices can be tailored to fit survey priority objectives.

In this paper we use some properties of block designs that are used in comparative experiments to derive rotation sampling designs for surveys. A block design is used to compare a number of "treatments," e.g., different drugs, different fertilizers, etc. The available experimental units, e.g., human volunteers, plots of agricultural land, etc., are divided up into "blocks" so that the units within a block are as similar as possible. The units in different blocks may then differ greatly. In the analysis of the data obtained from block experiment, the differences between the treatments are estimated from within-block comparisons. In an incomplete block design the number of units in a block is smaller than the number of treatments. This means that the estimates of the treatment differences must be adjusted for the differences between the blocks. An instructive review of the similarities between experimental design and sample survey methodologies is provided by Fienberg and Tanur (1985). They draw analogies between the design concepts in the two areas such as randomization/probability sampling, blocking/stratification and split-plots/clusters. Then they discuss the use of balanced incomplete block designs as a means of achieving a restricted randomization in sampling by reducing the "support" of a sampling plan (Chakrabarti (1963)). The support is the set of samples with positive selection probabilities. These authors then review analogies between the model-based analyses used in the two areas: while Model I or fixed effects linear models have seen several applications in sampling, Model II or random effects models are used rarely in the sampling literature (exceptions include Hartley and Rao (1978) and Fuller and Harter (1985)). Our paper is something of a synthesis in that linear models inspired by incomplete block designs are used to guide the search for optimal sampling designs. We use the framework of what are known as cyclic incomplete block designs to obtain alternatives to the symmetric designs used in sample surveys. Our designs reduce the response burden.

2. Rotation Sampling Designs

Many large-scale surveys are repeated periodically using a rotation sampling design. In such a design, a total sample is divided into b rotation groups. A rotation group can consist of a fixed number of primary sampling units (PSU's), segments within PSU's, or last stage units. Let rotation group j, where $j=1, 2, \ldots, b$, be interviewed in each of the k_j periods. The total number of periods in the design is t and $k_j \leq t$, for all j. Furthermore, let f be the constant fraction of the groups interviewed in one period that are also interviewed in the next period; the remaining fraction $(1-f)$ of the groups are replaced. The cyclical nature of our designs should be noted. In each period, r groups are interviewed, and this requires that a particular group returns to the sample.

The design used in the National Crime Survey (Fienberg (1980)), for example, has this type of overlap pattern (slightly modified to accommodate bounding of interviews). Other rotation patterns are used in the Current Population Survey carried out by the U.S. Bureau of the Census (1978) and in Statistics Canada's Labor Force Survey (Ghangurde (1982)).

A convenient representation of this design is given by the *incidence matrix* N, which has elements n_{ij}, $i=1, 2, \ldots, t$; $j=1, 2, \ldots, b$. The element $n_{ij}=1$ if group j is interviewed in period i, and is zero otherwise. The analogy with the construction of block designs, as used in comparative experiments, is that

b = the number of blocks,
t = the number of treatments,
r = the number of units which receive each particular treatment,
k_j = the number of units in block j, and
n_{ij} = the number of units in block j which receive treatment i.

The class considered here is an important subclass of the class of cyclic designs (see, e.g., John et al. (1972)).

3. Construction

A rotation design is completely specified by its incidence matrix N. The basic designs considered in this paper have an incidence matrix that is constructed as follows. Given b, r, and a "shift parameter" s:

1. Set the first r entries of the first row equal to 1 and the rest to 0. That is, set $n_{1j}=1$; $j=1, 2, \ldots, r$ and $n_{1j}=0$ for $j=r+1, \ldots, b$.
2. Construct the second row of N by shifting all entries in the first row s places to the right in a cyclical manner. That is, the $(p+s)$ th element of row 2 is set equal to the p th element of row 1, where $(p+s)$ is taken modulo b, and $p=1, 2, \ldots, b$.
3. Rows 3, 4, ... are constructed in exactly the same way from their directly preceding rows.
4. Eventually row 1 will occur in the $(t+1)$ th row. This last row is discarded to leave an incidence matrix with t rows and b columns. Then we see that the value of k_j is the number of ones in column j;$j=1, 2, \ldots, b$.

These types of designs were constructed for $b=1,2,\ldots,12$. For each value of b, designs were systematically obtained by varying r from 1 to b–1, and for each r by varying s from 1 to r. As an example, Table 1 gives the incidence matrix for $b=8$, $r=5$, and $s=2$, constructed as explained above. This design has $t=4$, $k_1=k_3=k_5=k_7=3$, and $k_2=k_4=k_6=k_8=2$. More realistic examples will be given in the following sections.

Table 1. Incidence Matrix for b=8, r=5 and s=2.

1	1	1	1	1	0	0	0
0	0	1	1	1	1	1	0
1	0	0	0	1	1	1	1
1	1	1	0	0	0	1	1

An additional parameter that is useful for describing a design is the "overlap," l, between rows. In Table 1 the overlap is $l=3$. It will be noted that $f=\frac{l}{r}$. As will be seen in the following, most designs have $k_j=k$, a constant, for all j. Designs for which the incidence matrices are the same (except for a permutation of their columns) are considered equivalent, since they correspond to a simple relabelling of the groups.

4. Classification of the Designs

In order to classify the rotation designs, we introduce a linear model for the characteristic, Y, that is measured in the survey. For $i=1, 2, \ldots, t$ and $j=1, 2, \ldots, b$, we assume

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \tag{4.1}$$

where

Y_{ij} = the observed value of the characteristic for group j in period i,
μ = the overall mean,
α_i = the effect of period i.
β_j = the effect of group j,
and
ε_{ij} is an independent random residual with mean zero and variance σ^2.

It will be noted that the model is written in terms of the group characteristic rather than in terms of the characteristics of the individual units within each group, because in block designs it is unusual to have more than one response from each treatment (period) in a block (group). In sampling where there is usually more than one unit in each period/group combination, the above model could be applied to the mean value of the characteristic in each group in each period. This simple model is adequate for our purpose of obtaining efficient designs for comparing periods. This point and others are taken up in Section 7. We assume that the main objective of a given survey is to obtain information on contrasts, i.e., comparisons, between the period effects. We are interested in estimating $\sum_{i=1}^{t} c_i \alpha_i$, where $\sum_{i=1}^{t} c_i = 0$. Such contrasts include all period-to-period changes, but exclude the mean response in each period. The above formulation is congruous with Wolter (1979) and also with Gurney and Daly (1965).

To obtain the classification of our designs we consider comparisons of neighbouring periods. The "1st neighbour " comparisons are $\alpha_1-\alpha_2$, $\alpha_2-\alpha_3$, ..., $\alpha_{t-1}-\alpha_t$, $\alpha_t-\alpha_1$, where here and in the following, the comparisons are taken cyclically. The "2nd neighbour" comparisons are $\alpha_1-\alpha_3$, $\alpha_2-\alpha_4$, ..., $\alpha_{t-2}-\alpha_t$, $\alpha_{t-1}-\alpha_1$, $\alpha_t-\alpha_2$. In a similar manner, 3 rd, 4 th, ..., neighbour comparisons can be defined for a suitable t. The estimates $\hat{\alpha}_i$ are taken to be the usual least squares estimates, adjusted for group effects.

The (model) variance of these comparisons is of interest and we define

$$v_n \sigma^2 = V(\hat{\alpha}_i - \hat{\alpha}_j), \qquad (4.2)$$

where $\hat{\alpha}_i$ and $\hat{\alpha}_j$ are the estimated effects of two n th neighbours, $n=1, 2, \ldots, m_t$ and

$$m_t = \frac{(t-1)}{2} \text{ if } t \text{ is odd and}$$

$$m_t = \frac{t}{2} \text{ if } t \text{ is even.} \qquad (4.3)$$

Some of these variances may be identical and we denote the number of distinct ones by d.

A useful class of designs has

$v_1 \leqslant v_2 \leqslant \cdots \leqslant v_{m_t}$.

This class is introduced initially, not only for its practical interest, but for its conceptual import and its simplicity. The properties of rotation designs are more clearly understood for these monotonic designs. Furthermore, these basic designs are building blocks for more complex designs, as will be described later.

These designs give more importance to comparisons between neighbours that are closer together. That is, assign the smallest variance to estimated changes between 1st neighbours, next smallest to changes between 2nd neighbours and so on. In particular, this means that period-to-period, e.g., month-to-month, changes will be estimated precisely, a desirable characteristic in practice.

Let

$v_1 < v_2 < \cdots < v_{d-1} < v_d = v_{d+1} = \cdots = v_{m_t}$.

We denote this class of designs by B_d^+. Note that there is a dual class B_d^-, say, with just the opposite property, namely,

$$v_1 > v_2 > \cdots > v_{d-1} > v_d = v_{d+1} = \cdots = v_{m_t}.$$

The latter class may be of interest, for example, if $m_t = \frac{t}{2}$ and mid-cycle changes are the main focus of attention. For instance, for annual cycles ($t=12$) semestral changes might need to be estimated precisely. It will be seen in the next sections, moreover, that general rotation designs can be constructed by "mixing" (adjoining/combining) basic designs in these simple, monotonic, classes.

Balanced designs form the class B_1^+ with one single value for v_n. It is worthwhile to remark that for a symmetric balanced incomplete block design, l corresponds to the usual design parameter λ, where λ is the number of times a pair of treatments occur together in the same block. Designs currently used in sample surveys can often be considered replicates of symmetric balanced incomplete block designs and can also be represented as cyclic designs. The rotation sampling designs, however, are cyclic designs which reduce respondent burden. It may be of interest to note that the properties of a cyclic incomplete block design can be considered in terms of a corresponding paired comparison design where each block of size k is thought of as being divided into $\frac{1}{2}k(k-1)$ blocks of size two. We do not pursue this correspondence here since not all pairwise comparisons are of interest. (See John (1966) for further details.)

Table 2 lists some of the more efficient rotation designs of the $B_d{}^+$ type, constructed using the method described earlier. In the table some designs have a serial number which is followed by a C. These designs are complements of their partners with the same serial number. A complement of a design with incidence matrix N is a design with incidence matrix $N_c = J - N$, where J is a $t \times b$ matrix of ones. Note that $d_c = d$, $r_c = b - r$, and $k_c = t - k$ relate the parameters of a design to those of its complement.

Table 2. Classification of Rotation Designs: $b=3$ to 12, $2 \leqslant r \leqslant b-1$

Serial No.	b	t	r	k*	l	d**
1	3	3	2	2	1	1*
2	4	4	2	2	1	2*R
3	4	4	3	3	2	1*
4	5	5	2	2	1	2*
4C	5	5	3	3	2	2*
5	5	5	4	4	3	1*
6	6	6	2	2	1	3*
6C	6	6	4	4	3	3*
7	6	6	3	3	2	3*R
8	6	3	4	2	2	1
9	6	6	5	5	4	1*
10	6	3	5	(2,3)	4	1
11	7	7	2	2	1	3*
11C	7	7	5	5	4	3*
12	7	7	3	3	2	3*
12C	7	7	4	4	3	3*
13	7	7	6	6	5	1*
14	8	8	2	2	1	4*
14C	8	8	6	6	5	4*
15	8	8	3	3	2	4*
15C	8	8	5	5	4	4*
16	8	8	4	4	3	4*R

(cont).

Table 2. (Cont.) Classification of Rotation Designs: b=3 to 12, 2 ⩽ r ⩾ b–1

Serial No.	*b*	*t*	*r*	*k**	*l*	*d***
17	8	4	4	2	2	2R
18	8	4	5	(2,3)	3	2
19	8	4	6	3	4	1
20	8	8	7	7	6	1*
21	8	4	7	(3,4)	6	1
22	9	9	2	2	1	4*
22C	9	9	7	7	6	4*
23	9	9	3	3	2	4*
23C	9	9	6	6	5	4*
24	9	9	4	4	3	4*
24C	9	9	5	5	4	4*
25	9	3	6	2	3	1
26	9	3	7	(2,3)	5	1
27	9	9	8	8	7	1*
28	9	3	8	(2,3)	7	1
29	10	10	2	2	1	5*
29C	10	10	8	8	7	5*
30	10	10	3	3	2	5*
30C	10	10	7	7	6	5*
31	10	10	4	4	3	5*
31C	10	10	6	6	5	5*
32	10	5	4	2	2	2
32C	10	5	6	3	4	2
33	10	10	5	5	4	5*R
34	10	5	5	(2,3)	3	2R
35	10	5	7	(3,4)	5	2
36	10	5	8	4	6	1
37	10	10	9	9	8	1*
38	10	5	9	(4,5)	8	1
39	11	11	2	2	1	5*
39C	11	11	9	9	8	5*
40	11	11	3	3	2	5*
40C	11	11	8	8	7	5*
41	11	11	4	4	3	5*
41C	11	11	7	7	6	5*
42	11	11	5	5	4	5*
42C	11	11	6	6	5	5*
43	11	11	10	10	9	1*
44	12	12	2	2	1	6*
44C	12	12	10	10	9	6*
45	12	12	3	3	2	6*
45C	12	12	9	9	8	6*

(cont).

Table 2. (Cont.)

Serial No.	*b*	*t*	*r*	*k**	*l*	*d***
46	12	12	4	4	3	6*
46C	12	12	8	8	7	6*
47	12	6	4	2	2	3
47C	12	6	8	4	6	3
48	12	12	5	5	4	6*
48C	12	12	7	7	6	6*
49	12	6	5	(2,3)	3	3
49C	12	6	7	(3,4)	5	3
50	12	12	6	6	5	6*R
51	12	6	6	3	4	3R
52	12	4	6	2	3	2R
53	12	4	7	(2,3)	4	2
54	12	4	8	(2,3)	5	2
55	12	3	8	2	4	1
56	12	6	9	(4,5)	7	3
57	12	4	9	3	6	1
58	12	3	9	(2,3)	6	1
59	12	6	10	5	8	1
60	12	4	10	(3,4)	8	1
61	12	3	10	(2,3)	8	1
62	12	12	11	11	10	1*
63	12	6	11	(5,6)	10	1
64	12	4	11	(3,4)	10	1
65	12	3	11	(2,3)	10	1

* Values of k_j, $j=1, 2, \ldots, b$. Either $k_j=k$, all j, or $k_j=k$ or k', $k'=k+1$.
** Value of d in B_d^+.

Also, in Table 2, some designs have values of d which are written as d^*. These designs are symmetric and are such that $b=t$, $r=k$ and N can be written as a symmetric matrix, after possibly permuting its columns. Some designs are reflexive in the sense that taking their complements does not alter the designs themselves. These designs are indicated by an *R* in Table 2. We further denote by P_t the symmetric balanced design with $r=k=t-1$. (Note that all symmetric designs with a given $b=t$ have $d=m_t$ as in (4.3).)

Designs which have a block size of 1, or are complements which have $r=1$, have been omitted from Table 2. Such designs are not considered useful for the estimation of contrasts, i.e., changes. Moreover, for each of the designs in Table 2 the variances v_n were obtained and are considered in Section 6.

5. Properties of the Designs

Other rotation designs can be obtained by adjoining two or more basic rotation designs, e.g., those listed in Table 2. A design D_1 is adjoined to another D_2 by taking the b_1 columns of D_1 and putting them alongside the b_2 columns of D_2. Of course, both D_1 and D_2 must have the same value of t.

Some of the designs constructed in this way had an incidence matrix which was obtained by adjoining a $t \times t$ identity matrix I_t, to an incidence matrix which had two or more 1's in each column. For example, a design for $b=8$, $t=4$, $r=3$, $k_j=2$ or 1, $j=1, 2, \ldots, b$, is obtained by adjoining I_4 to the incidence matrix for the second design in Table 2, and suitably permuting the columns. Designs with an identity matrix component have been excluded from Table 2.

Table 3 gives a list of the designs in Table 2, which are, in fact, obtained by adjoining other designs in Table 2. Clearly, other designs not in Table 2 could be constructed by adjoining designs that are in Table 2. This adjoining is, of course, subject to the final design being a rotation design.

Table 3. Designs Obtained by Adjoining Two or Three Other Designs

Serial No. in Table 2	Serial Nos. of adjoined designs
17	2,2
18	2,3
52	2,2,2
53	2,2,3
54	2,3,3
32	4,4
34	4,4C
32C	4C,4C
35	4C,5
47	6,6
49	6,7
49C	7,6C
47C	6C,6C
56	6C,9
51	7,7

Clearly, a design obtained by adjoining two balanced (B_1^+) designs is also balanced. Further, adjoining a B_2^+ and a B_1^+ design leads to another B_2^+ design. The effect of adjoining one or more B_1^+ designs to a B_2^+ design is to make the variances, v_1 and v_2, more similar in size (see Section 6).

A different operation consists of combining (vertically) designs with a common b, r, and l to obtain larger rotation designs. In fact, we note that each design of type P_t, for non-prime $t=h_1h_2$, may be formed by combining its "component" design with $t=h_1$ (h_2 times), and similarly for $t_2=h_2$ (h_1 times). Many of these components have blocks of size 1 and are not given in Table 2. Table 4 presents the components of designs P_t, for $4 \leq t \leq 10$.

Table 4. Subclass P_t of Symmetric Balanced Designs, $4 \leq t \leq 10$, and their components[1]

	b	t	r	k	l	d
$P_4 = D_1 \oplus D_1$:	4	4	3	3	2	1
Component: D_1	4	2	3	(1,2)	2	1
$P_6 = D_2 \oplus D_2$						
$= D_3 \oplus D_3 \oplus D_3$:	6	6	5	5	4	1
Components: D_2	6	3	5	(2,3)	4	1
D_3	6	2	5	(1,2)	4	1
$P_8 = D_4 \oplus D_4$						
$= D_5 \oplus D_5 \oplus D_5 \oplus D_5$:	8	8	7	7	6	1
Components: D_4	8	4	7	(3,4)	6	1
D_5	8	2	7	(1,2)	6	1
$P_9 = D_6 \oplus D_6 \oplus D_6$	9	9	8	8	7	1
Component: D_6	9	3	8	(2,3)	7	1
$P_{10} = D_7 \oplus D_7$						
$= D_8 \oplus D_8 \oplus D_8 \oplus D_8 \oplus D_8$:	10	10	9	9	8	1
Components: D_7	10	5	9	(4,5)	8	1
D_8	10	2	9	(1,2)	8	1

[1] The operation of combining (vertically) two designs, D_1 and D_2, say, is denoted by $D_1 \oplus D_2$.

Of course, in any real continuing survey t is not fixed and the designs would evolve over time by adjoining and combining designs. The effects of adjoining and combining designs, as described above, would then be of particular interest. For example, a rotation pattern similar to that used in the U.S. National Crime Survey (NCS) can be generated by successive adjoining and combining operations. By combining (6×12) matrix $C = (A|B)$ to $(O|A)$ one derives the (NCS) pattern displayed in Table 5. Note that the (6×12) matrix $C = (A|B)$ is of the elementary type described in Sections 2 and 3. The pattern employed in the U.S. Current Population Survey can be generated by more complex mixtures of the elementary rotation designs considered in Table 2.

Table 5. NCS Rotation Pattern

t	Rotation groups (Panels) (A)						(B)					
1	1	1	1	1	1	1	0	0	0	0	0	0
2	0	1	1	1	1	1	1	0	0	0	0	0
3	0	0	1	1	1	1	1	1	0	0	0	0
4	0	0	0	1	1	1	1	1	1	0	0	0
5	0	0	0	0	1	1	1	1	1	1	0	0
6	0	0	0	0	0	1	1	1	1	1	1	0
⋮	(0) 6×6						(A)					

6. Variances of Estimated Period Changes

The values of v_n, where $v_n\sigma^2 = V(\hat{\alpha}_i - \hat{\alpha}_j)$ are given in Table 6 for $d=1, 2$ and 3 only, to save space, the remaining ones being available from the authors. The corresponding efficiency factors for each variance are also given. The efficiency factor E_n, say, is defined as

$$E_n = \frac{V(\hat{\alpha}_i - \hat{\alpha}_j) \text{ in a saturated design}}{V(\hat{\alpha}_i - \hat{\alpha}_j) \text{ in the design under consideration}}$$

$$= \frac{2}{r\, v_n}.$$

Here, a saturated design is one in which every group is interviewed in every period. This definition is the one used in the design of comparative experiments. The saturated design of such experiments is the randomized (complete) block design, with r blocks, each of size t. Costs usually prohibit the use of a saturated design in a survey. Nevertheless the saturated design does provide a useful yardstick for measuring the efficiency of other designs. It would be possible to compare the average value of v_n with $\frac{2}{r}$, but this would conceal the true range of the efficiencies in a given design.

For comparison purposes, the total size $n = r \times t$ of the designs are also included in Table 6.

Table 6. Values of the Variances (v_n) and Efficiencies for Designs with d = 1, 2 and 3

Serial No.	b	t	r	n^1	Variances (v_n)			Efficiencies ($\frac{2}{rv_n}\times 100\%$)		
1	3	3	2	6	1.3333			75		
2	4	4	2	8	1.5000	2.0000		67	50	
3	4	4	3	12	0.7500			89		
4	5	5	2	10	1.6000	2.4000		63	42	
4C	5	5	3	15	0.7636	0.8727		87	76	
5	5	5	4	20	0.5333			94		
6	6	6	2	12	1.6667	2.6667	3.0000	60	37	33
6C	6	6	4	24	0.5359	0.5744	0.5769	93	87	87
7	6	6	3	18	0.7833	0.9333	1.0500	85	71	63
8	6	3	4	12	0.6667			75		
9	6	6	5	30	0.4167			96		
10	6	3	5	15	0.4444			90		
11	7	7	2	14	1.7143	2.8571	3.4286	58	35	29
11C	7	7	5	35	0.4174	0.4356	0.4364	96	92	92
12	7	7	3	21	0.7944	0.9826	1.1498	84	68	58
12C	7	7	4	28	0.5396	0.5881	0.6336	93	85	79
13	7	7	6	42	0.3429			97		
17	8	4	4	16	0.7500	1.0000		67	50	
18	8	4	5	20	0.4870	0.5455		82	73	
19	8	4	6	24	0.3750			89		
20	8	8	7	56	0.2917			98		
21	8	4	7	28	0.3000			95		
25	9	3	6	18	0.4444			75		
26	9	3	7	21	0.3333			86		
27	9	9	8	72	0.2540			98		
28	9	3	8	24	0.2667			94		
32	10	5	4	20	0.8000	1.2000		63	42	
32C	10	5	6	30	0.3818	0.4364		87	76	
34	10	5	5	25	0.5053	0.6316		79	63	
35	10	5	7	35	0.3126	0.3297		91	87	
36	10	5	8	40	0.2667			94		
37	10	10	9	90	0.2250			99		
38	10	5	9	45	0.2286			97		
43	11	11	10	110	0.2020			99		
47	12	6	4	24	0.8333	1.3333	1.5000	60	37	33
47C	12	6	8	48	0.2679	0.2872	0.2885	93	87	87

(cont.)

Table 6. (Cont).

Serial No.	b	t	r	n[1]	Variances (v_n)			Efficiencies ($\frac{2}{rv_n} \times 100\%$)		
49	12	6	5	30	0.5189	0.6838	0.7582	77	58	52
49C	12	6	7	42	0.3159	0.3515	0.3712	90	81	77
51	12	6	6	36	0.3917	0.4667	0.5250	85	71	63
52	12	4	6	24	0.5000	0.6667		67	50	
53	12	4	7	28	0.3643	0.4286		78	67	
54	12	4	8	32	0.2943	0.3158		85	79	
55	12	3	8	24	0.3333			75		
56	12	6	9	54	0.2341	0.2412	0.2414	95	92	92
57	12	4	9	36	0.2500			89		
58	12	3	9	27	0.2667			83		
59	12	6	10	60	0.2083			96		
60	12	4	10	40	0.2143			93		
61	12	3	10	30	0.2222			90		
62	12	12	11	122	0.1833			99		
63	12	6	11	66	0.1852			98		
64	12	4	11	44	0.1875			97		
65	12	3	11	33	0.1905			95		

[1] Total sample size = $n = rt$.

Adjoining x copies of a design reduces the value of v_n in that design by a factor x, but does not change E_n. For example, the design with serial number 8 can be constructed by adjoining two copies of design 1. Adjoining copies of a design of type B_2^+ to a design of type B_1^+, reduces the relative differences between the values of v_n. For example, adjoining designs 4C and 5 produces design 35. In design 4 C the ratio of the two variances is 1:1.5 but in design 35 is 1:1.05.

Since taking the complement of a design gives $r_c=b-r$, without changing b, the complement of a highly efficient design will have low efficiency. By observing the values of the efficiencies in Table 6, it is clear that, while most designs have high efficiency, there are some with particularly low efficiencies. The low efficiencies occur, as one might expect, for designs where the k_j values equal 2. In other words, all groups must be in the sample more than twice to achieve a reasonable efficiency.

The efficiencies allow a useful design to be chosen out of a number of competitors. For example, if t=6 and b=12 then design 49 is much more efficient than design 47 and only uses 6 more groups. Of course, design 47C is even more efficient than design 49 but requires 18 more groups.

7. Discussion

Some alternatives to the simple model (4.1) may be considered. First, it may be more realistic to consider the effects of individual groups being random rather than fixed. This would complicate the theory in Sections 4–6 but is not likely to substantially change the overall results.

A second alternative would be to model, individually, each unit within a group. The model for the measurement on the q *th* unit in the j th group in the i th period would then be

$$Y_{ijq} = \mu+\alpha_i+\beta_j+\gamma_{ij}+\varepsilon_{ijq}, \qquad (7.1)$$

where μ, α_i, β_j are defined as in Secton 4; γ_{ij} is the period-by-group interaction effect; and ε_{ijq} is the usual error term. This type of interaction is manifested in practice via what is called the rotation group bias (see, e.g, Bailar (1975)). This bias is related to the number of times different units have been interviewed previously. The presence of an interaction term will contaminate published period-to-period changes, i.e., estimates of the form $\hat{Y}_i - \hat{Y}_{i'}$, which are of more direct interest than contrast estimates $\hat{\alpha}_i - \hat{\alpha}_{i'}$. Clearly, $\hat{Y}_i - \hat{Y}_i'$ differences will involve the interaction parameters γ_{ij}.

Another model-based approach employs time series models to estimate the current mean response. References for this approach are Blight and Scott (1973), Scott and Smith (1974), and Jones (1980).

We have discussed in Section 4 the role of the monotonic classes B_d^+ and B_d^-; in particular, Table 2 is concerned only with B_d^+. We have also examined, however, other diverse variance patterns that are not reported here. Of particular interest is an investigation of the duals or "mirror images" in B_d^- of designs in the class B_d^+. A summary for t=5, d=2 and the designs in Table 2 is presented in Table 7.

Table 7. Mirror Images of Designs for t=5 and = d = 2

Serial No.	Image	*b*	*t*	*r*	*k*	*l*
4	4I	5	5	2	2	0
4C	4CI	5	5	3	3	1
32	32I	10	5	4	2	0
32C	32CI	10	5	6	3	2
34	34I	10	5	5	(2,3)	1
35	35I	10	5	7	(3,4)	4

It is worthwhile to remark that the approach in Sections 4–6 has completely ignored the randomization imposed by the design. In particular, all variances are considered with respect to the model distribution. Denoting such variances by V_m and expectations over the sampling design, p, say, by E_p, one may also consider the criterion E_pV_m. In this context, Bellhouse (1984) has derived optimal treatment assignments for certain subclasses of treatment contrasts. Other references where complex sample design effects have been considered include Nathan and Holt (1980) and Holt, et al. (1980).

8. References

Bailar, B.A. (1975): The Effects of Rotation Group Bias on Estimates from Panel Surveys. Journal of the American Statistical Association, 70, pp. 23–30.

Bellhouse, D.R. (1984): Optimal Randomization for Experiments in Which Autocorrelation is Present. Biometrika, 71, pp. 155–160.

Blight, B.J.N. and Scott, A.J. (1973): A Stochastic Model for Repeated Surveys. Journal of the Royal Statistical Society, B, 35, pp. 61–66.

Chakrabarti, M.C. (1963): On the Use of

Incidence Matrices of Designs in Sampling from Finite Populations. Journal of the Indian Statistical Association, 1, pp. 78–85.

Cochran, W.G. (1977): Sampling Techniques. Third Edition, Wiley, New York.

Eckler, A.R. (1955): Rotation Sampling. Annals of Mathematical Statistics, 26, pp. 664–685.

Fienberg, S.E. (1980): The Measurement of Crime Victimization: Prospectus for Panel Analysis of a Panel Survey. The Statistician, 29, pp. 313–350.

Fienberg,S.E. and Tanur, J.M. (1985): A Long and Honorable Tradition: Intertwining Concepts and Constructs in Experimental Design and Sample Surveys. Proceedings of the International Statistical Institute Meetings in Amsterdam, 45th Session, Vol. L1.

Fuller, W.A. and Harter, R. (1985): The Multivariate Components of Variance Model for Small Area Estimation. Paper presented at the Small Area Symposium, Ottawa, Canada.

Ghangurde, P.D. (1982): Rotation Group Bias in the LFS Estimates. Paper presented at the meeting of the American Statistical Association in Cincinnati.

Ghangurde, P.D. and Rao, J.N.K. (1969): Some Results on Sampling over Two Occasions. Sankhya, 31, pp. 463–472.

Gurney, M. and Daly, J.F. (1965): A Multivariate Approach to Estimation in Periodic Sample Surveys. American Statistical Association, Proceedings of the Social Statistics Section, pp. 242–257.

Hartley, H.O. and Rao, J.N.K. (1978): Estimation of Nonsampling Variance Components in Sample Surveys. In N.K. Namboodiri (Ed.), Survey Sampling and Measurement. Academic Press, New York.

Holt, D., Smith, T.M.F., and Winter, P.D. (1980): Regression Analysis of Data from Complex Surveys. Journal of the Royal Statistical Society, A, 143, pp. 474–487.

Jain, A.K. (1981): Design of a Rotation Scheme for a Stratified Multi-Stage Sample. Journal of Statistical Planning and Inference, 5, pp. 57–69.

John, J.A. (1966): Cyclic Incomplete Block Designs. Journal of the Royal Statistical Society, B, 28, pp. 345–360.

John, J.A., Wolock, F.W., and David, H.A. (1972): Cyclic Designs. National Bureau of Standards, Applied Mathematics Series, 62.

Jones, R.G. (1980): Best Linear Unbiased Estimators for Repeated Surveys. Journal of the Royal Statistical Society, B, 42, pp. 221–226.

Nathan, G. and Holt, D. (1980): The Effect of Survey Design on Regression Analysis. Journal of the Royal Statistical Society, B, 42, pp. 377–386.

Paterson, H.D. (1950): Sampling on Successive Occasions with Partial Replacements of Units. Journal of the Royal Statistical Society, B, 12, pp. 241–255.

Scott, A.J.L and Smith, T.M.F. (1974): Analysis of Repeated Surveys Using Time Series Methods. Journal of the American Statistical Association, 69, pp. 674–678.

U.S. Bureau of the Census (1978): The Current Population Survey: Design and Methodology. Technical paper 40, U.S. Government Printing Office, Washington, D.C.

Wolter, K.M. (1979): Composite Estimation in Finite Populations. Journal of the American Statistical Association, 74, pp. 604–613.

Received December 1985
Revised April 1987