NATIONAL HOUSEHOLD SURVEY CAPABILITY PROGRAMME

Sampling Frames and Sample Designs for Integrated Household Survey Programmes



DP/UN/INT-84-014/5E

NATIONAL HOUSEHOLD SURVEY CAPABILITY PROGRAMME

Sampling Frames and Sample Designs for Integrated Household Survey Programmes Preliminary version

.

UNITED NATIONS DEPARTMENT OF TECHNICAL CO-OPERATION FOR DEVEOPMENT and STATISTICAL OFFICE

New York, 1986

PREFACE

This study is one of a series of publications designed to assist countries in planning and implementing household surveys in the context of the National Household Survey Capability Programme. The United Nations revised <u>Handbook of Household Surveys</u>* is the basic document in the series. The <u>Handbook</u> reviews issues in survey content, design and operations and provides technical information and guidance at a relatively general level to national statistical organizations charged with carrying out household survey programmes. In addition to the <u>Handbook</u>, a number of studies have been undertaken to provide reviews of issues and procedures in specific areas of household survey methodology and operations and in selected subject areas. The major emphasis of the series is that of continuing programmes of household surveys.

The topics covered in this study are (a) the development and maintenance of sampling frames; and (b) sample designs for an integrated programme of household surveys with particular attention to the use of master samples. Both topics are of central importance in the undertaking of integrated programmes of household surveys. There are many excellent text books which deal with the theoretical and practical aspects of sample design for a single household survey but they do not discuss sampling and related issues that arise in designing a programme of surveys intended to provide data on several topics over an extended period of time. This study attempts to fill this gap.

In the preparation of this document, the United Nations was assisted by Mr. Thomas B. Jabine serving as a consultant. The study was based on a detailed outline developed by the United Nations Statistical Office after extensive consultations with a number of survey statisticians from both developed and developing countries. The document is being issued in a preliminary version to obtain further comments and feedback from as many readers and users as possible prior to its publication in final form.

*Studies in Methods, Series F, No. 31 (ST.ESA.STAT.SER.F/31).

Contents

I.	INTF	RODUC	TIO	Ν.	•	•	• •	•	•	•	•	•	•	•	•	•	•	•	•	•	•	1
	Α.	Audi	enc	e,	sco	pe	ar	nd	ge	ne	ra	1 8	app	pro	ac	h	•	•	•	•	•	2
	в.	Orga	niza	ati	on	of	th	nis	đ	oc	um	ent	t	•	•	•	•	•	•	•	•	3
II.	BASI	c co	DNCE	PTS	AN	DI	DEF	PIN	IT	10	NS	•	•	•	•	•	•	•	•	•	•	5
	A.	Inte	gra	tio	n i	n a	a s	ys	te	m	of	h¢	ous	eh	101	d	su	rv	ey	s	•	5
	в.	Surv	vey i	uni	ts	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	7
	с.	Mult	ista	age	sa	mp:	lin	ıg	•	•	•	•	•	•	•	•	•	•	•	•	•	9
	D.	Samp	lin	g f	ram	es	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
	Ε.	Mast	er	sam	ple			•	•	•	•	•	•	•	•	•	•	•	•	•	•	11
	F.	Opti	mum	su	rve	у	des	sig	n	•	•	•	•	•	•	•	•	•	•	•	•	11
III.	DESI	GNS	FOR	IN	FEG	RA'	FEI	H	OU	SE	HO	LD	รเ	JRV	/EY	ſ	PRO	GR	AM	IME	S	14
	A.	Gene	eral	de	sig	n i	req	ļui	re	me	nt	s	•	•	•	•	•	•	•	•	•	14
		1.	Lon	g-ra rdi	ang	e j i o	pla n w	inn 71 t	in h	g th	•	•	• • • • •	lat	.ic	• • n	•	•	•	•	•	15
		2.	CCC.	ensi	us							<u> </u>	242					_				16
		3.	Fle	rib		t.v	•••															17
		4	Use	of	pr	oĥa	abi	1 i	tv	s	ami	o] i	ind	ד.	-		-				-	18
		5.	Doc	ume	nta	tio	on	•	•	•	•	•	•	•	•	•	•	•	•	•	•	18
	в.	Data	re	qui	rem	ent	ts	•	•	•	•	•	•	•	•	•	•	•	•	•	•	20
		1	ന്ന	ice	fo	r '1	hou	مە	ho	ıa	q		VA	79					_		_	21
		2	Sper	nia	1 4		i an	130	00 00	n i	7 01		nt o	2	•	•	•	•	•	•	•	23
		2.	Gro	uni	1 U	+ ~1			ui ui	4 1. + h	- C/	i m'	i 1 :	, 	•	•	•	•	•	•	•	23
		5.	C C	equ	ire	mer	nts	3.	• •	•	•	•	•	•	•	•	•	•	•	•	•	26
	с.	Oper	ati	ng	env	ir	onn	nen	t	an	đ	coi	nst	:ra	ain	its	8	•	•	•	•	29
		1	۵dm	ini	et r	at :	ive	. a	tr		+ n	re	of		ווסי	nt	r v	,	_	_	_	31
		2.	Cha	rac	ter	is	tic	s	of	t	he	t	arq	get	:				•	•	•	
		_	_ P	opu.	lat	101	n.		•	•	•	•	•	٠	٠	٠	•	٠	٠	•	٠	33
		3.	ACC	ess	to	S	amp	bie bie	u	ní	ts	٠	•	•	• .	•	•	•	•	•	٠	54
		4.	Ava.	ila	bil	ity	уc	DE	sa	mp	11	ng	Ma	ate	eri	a]	S	•	٠	•	•	35
		5.	Fie	TQ :	sta	ff		•	٠	•	. :	•	•	٠	٠	٠	٠	•	٠	•	٠	36
		6.	Data	ар	roc	ess	sin	ıg	ca	pa	bi	11	ti€	98	•	٠	٠	٠	•	٠	•	36
		7.	Tec	hni	cal	aı	nd	ma	na	ge	ri	al	st	:af	f	•	•	•	•	•	•	37

	D.	Gen	era	1 (cla	s	ses	s c)f	d	es	i	gn	S	f	or	j	Ĺn	te	gi	rat	:ec	E			
		h	ous	eho	bld	l s	sur	ve	ey.	р	ro	g	ra	mn	e	S	•	•	•	•	•	•	•	•	•	38
		-				•	_	_	_		- 1	_	_	1	· •				- -		_	•				
		⊥.	De	sig	JN	Α:		a	S	1 n	gı	.e	R	L DI	Lτ	1-	sι	d	Je	C						
		•	_	sur	:ve	Y.	•	•	•	•	•		•	•	•	•			•	•,	•	•	٠	٠	•	39
		2.	De	sig	jn	в:		t٧	10	0	r	M	or	e	S	lN	g	Le	- 9	sur	с]€	ect				
				sur	ve	٩Y	5	٠	٠	٠	•		•	٠	•	•	•	•	•	٠	٠	٠	٠	٠	•	43
		3.	Ch	005	sin	g	an	l a	p	pr	op)r:	ia	ite	9 (de	si	ig	n	•	•	٠	•	٠	•	47
IV.	SAM	PLIN	G F	RAN	1ES	;	•	•	•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	49
	Α.	Bas	ic	cor	nsi	de	era	ıti	0	ns	i	n	t	:he		ch	oi	ic	е	of	E					
			sam	pli	ing	j 1	Ēra	me	s	•	•		•	•	•	•			•	•	•	•	•	•	•	50
		1	Tn	+ 01	ođe	5																				50
		1. 2	- 111 - 111			:u 	u3 :		•	•	•		•	•	•	•	•		•	•	•	•	•	•	•	50
		2.		ame	: u	111	LLS)	•	•	•		•	•	•	•	•	•	•	•	٠	٠	•	•	•	53
		3.		ver	ag	je	•	•	٠	٠	•		•	•	•	•	•		•	٠	•	٠	•	•	•	23
		4.	Me	a 1a		•	•	٠	٠	٠	•		•	٠	٠	•	•	•	•	٠	•	٠	٠	٠	•	54
		5.	CO	nte	ent		•	•	•	:	:		•	•	•	•	•	•	•	٠	•	٠	•	•	•	54
		6.	Au	X1]	11a	rj	γπ	at	.eı	[1	aı	S		•	٠	•	•	•	•	٠	•	٠	•	•	•	55
	в.	Fra	me	uni	its	; a	and	l r	u.	le	s	03	£	as	SS	oc	ia	at	ic	n	•	•	•	•	•	57
	с.	Des	ira	ble	e p	orc	ope	rt	i	es	o	f	f	ra	m	es		•	•		•				•	63
		1	011	ali	i + v	- 1	ام.	at	· •	a	nr	· ^ 1	ne	.rt	- i .	6 C										64
		2	Ēf	fic	l L y l i p	- L	. CI V-	.uc	12	u a+	PL		pc nr			c5 rt	i		•	•	•	•	•	•	•	70
		2.	<u>Co</u>	c+.	-r0	1 :	-1 -+-		- I (ro	ne			05		1 0	ц	-0		•	•	•	•	•	•	70
		J.	CU	3 L - mm -	- T C	, . T C	aue	u	P.		Þe		ιı	. 63	•	•	•	•	•	•	•	٠	•	٠	•	70
		4.	Su	uunc	i L y		•	•	•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	19
	D.	Dev	elo	pme	ent	. a	and	lπ	ia:	in	te	in	an	ice	e (of	ā	3	ma	ist	:ei	5				
		S	amp	<u></u> lir	ıg	fı	саπ	e	•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	79
		1.	De	+ 0 7	mi	n 2	at i	07		٦f	0		ne	ra	1	•	h-	i۵	c+	·i、	200	2				
		- •	20	and		+ 2		<u> </u>			9				• -	Ŭ	×.) C								81
		2	та	ont	- i f	i z	- a +	10	יני	• >	n.d		• • 1)		•	• + =	i .	• • •	•	, ÷	•	٠	•	•	•	01
		2.	TU	2110	- <u>-</u> <u>-</u> 	- 1 (10			ם יור	+ 0					aı	ТС	<i>.</i>	C	, ,						95
		2	Do	avc	117	ar	710	. T	 	2u ov	່ວ	ı h	•		•••			•	•	•	•	•	•	•	•	00
		J.	שכ		510	- 112 	50		. E	ey ?	0	and al		a.	1	er	E	56	10	- 9 			•	•	•	09
		4.	PL	epa a.	11a					a 2	5	CI L	ne -		ידו	e	гc)[T	. 11 1		a	L .			07
		F	De	aev	/ei	0	ome	int	: (30	C	.n.	e 	M2 -		•	•		•		•	٠	•	•	٠	97
		э.	De	vei the	.op M	enne I S I	2 N C 7	. C	Ξ	a	P	ιč	an	I	0	Ľ	up	pa	aτ	. 11	19					100
		6	ъđ	ant	 +	i c	.ne	•		- '	10	C	•	+ 1	•	. •	• *	, 10	• = 1	•	•	•	•	•	•	100
		••	ΠQ	cir	cu	ms	sta	inc	e	S	<u>те</u>		•	•	•	•	т с -		•	•	•	•		•		104
	F	Sac	ond	aru	, .	217	ומי	in		F	rə	m		. 1	C	C F	с)	•								108
	•		Gild	ur J	6	CI II	דקי	- 1)	.9	L	гd	.1119	3		5.	5 F	3/	,	•	•	•	•	•	•	•	100
		1.	SS	Fs	wi	tł	n a	re	ea	u	ni	t	S	•	•			•	•	•		•	•	•	•	108
		2.	SS.	Fs	wi	tł	ו ו	is	;t	u	ni	t	5	•	•	•	•			•	•	•	•	•	•	112
		3.	Di	scu	ISS	ic	on	an	d	S	uπ	ma	ar	Y	•	•			•	•	•	•	•	•	•	119

Table of Contents (cont'd)

ł

V.	MAS	rer s	SAMPI	LES	•	• •	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	123
	A.	What	: is	a n	as	ter	sa	mp	le	?	•	•	•	•	•	•	•	•	•	•	•	123
		1. 2.	The Defi	fir init	st io	ma: n ai	ste nd	er ke	sa Y	mp fe)le at	ur	es		•f	• a	•	•	•	•	•	123
			ma	aste	er	samj	ple)	•	•	•	•	•	•	•	•	•	•	•	•	•	124
	в.	Pros	s and	d co	ns	of	us	sin	g	ma	st	er	· 5	am	pl	es	5	•	•	•	•	128
		1.	Adva	anta	ıge	s.	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	128
		2.	Limj	itat	io	ns	•	•	•	•	•	•	•	•	•	•	•	٠	•	•	•	133
		3.	Sumn	nary	,	• •	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	137
	c.	Use	ofn	nast	er	sar	npl	.e	pr	in	ci	pl	es	s i	n							
		ጠሀ	ıltir	roun	d	sur	vey	/s	•	•	•	•	•	•	•	•	•	•	•	•	•	139
		1.	Some	e ex	am	ple	s		•								•	•			•	139
		2.	Disc	cuss	io	n o	Ēċ	les	ia	'n	is	su	ies	3								143
		-	a.	Ove	r 1.	ap l	bet	we	en	r	ou	nd	s									144
			h.	Sta	ae	ຮັດ	fs	am	ומ	in	σ			-	•		-	-				145
			c.	Use	· · ·	fs	e 1 f	w	ei	αh	it i	na	i s	am		ės			·		•	147
			а.	Eyh		sti	on on	. "	- -	· э . Απ	1 מו		່າກ	ni t				•	•	•	•	149
			<u> </u>	Dur		ion	on f	:	90		121		~ .			•	•	•	•	•	•	151
			C •	Dur		1011	01	. u		•	•	•	•	•	•	•	•	•	•	•	•	1 7 1
	D.	Use	ofn	nast	er	sai	mb]	les	f	or	m	u 1	ti	pl	e	su	ľ	/ey	'S	•	•	155
		1	Broa	ad o	hi	ect.	ive	g			_						_	_		_		155
		2	Case	a et	nd.	v i	111	+	• • • =	+ i	•	•	•	•	•	•	•	•	•	•	•	160
		2.	Come	- 30 	nu.	y 1.	1 I L ~ +	ist do		(L) (7)		10	•		•	•	•	•	•	•	•	162
		э.	Some		. 1	LLAI		ue .ho		. gu a	 	33	u c	: 0 . a t	•	•	•	• • • 1	:	•	•	102
			d .	поw	/ <u>1</u>	arg	e 2	sno	uı	.u	CI	e	1110	ISC	.er	3	a	цЪт	e			162
			L	- D	er L	•	•		•	•		•	•	•	• -	•		•			•	103
			D.	15		e u	se	OI	Ľ	eţ	11	ca	101	. On		es	11	ac)Te	: (•	100
			c.	Upo	ac	ing	•	•	•	•	•	•	•	٠	•	•	•	•	•	•	•	108
	Ε.	Spec	cial	toŗ	oic	s r	elā	ıti	ng	, t	0	ma	st	:er	9	an	ip]	les	1	•	•	172
		1.	Usin	ng a	n m	ast	er	sa	a m	le	e i	n	co	mb	in	at	ic	n				
			wi	ith	ot	her	sa	mp	lē	s					•	•	•	•				172
		2.	Can	mas	te	r sa	amc) le	s	fo	r	ho	us	eh	01	d	su	irv	eγ	's		
		-•	be	e us	ed	fo	r a	ar	ic	:u1	tu	ra	1	su	rv	ev	'sî	>	. 1			174
		3.	Trea	atme	nt	of	sr	ber	ia	1	חמ	יוס	1a	iti	on	 	r	our	s	-	•	176
		4	Oual	li+u	,	5.511	rar	n C P					u			. 3			-		•	177
		70	Yuu			554				•	•	•	•	•	•	•	•	•	•	•	•	±,,
VI.	SUMN	MARY	AND	CON	ICL	USI	ONS	5	•	•	•	•	•	•	•	•	•	•	•	•	•	179
	Α.	Summ	nary	•	•	•••	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	179

	в.	Re	CO	mm	en	da	ti	or	s	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		182
		1. 2.	1	Ch De De	oi si si	.ce gn	o o i o	f f	an a a	na ma	ve st	er er	11 9	. I am	HS [q]	SP lin le	d€ ig	esi fr	.gr an	n ne	•	•	•	•		182 182 183
		4.		De	ve fr	lο aπ	pm	er (it SS	an Fs	d)	us fc	e or	of ma	st	sec cer	or s	ida san	iry np]	, s le	san ur	np] nit	ir :s	ŋġ	•	184
		5.	1	Ge	ne	ra	1	cc	ns	id	ler	at	ic	na	•	•	•	•	•	•		•	•	I	•	185
ANNEX	I:	CA	SE	-s	TU	IDI	ES		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		186
	Aust	ra	1i	a	•	•					•	•	•			•		•				•				187
	Bots	ະພາ	na	-				-	-			-		-		-		•		-	•	-				194
	Ethi	on	ia		•	•	•	•	•	•	•															198
	Indi	i a			•		•		•			•									•					203
	Jord	lan	•	•	•		•	•	•			•	•			•	•	•	•	•						207
	More		~	•	•	•	•	•	•	•	•	•	•	•	•		•	•								212
	Nige		2	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		215
	RI 90	44 25 T(а \ -	• >h	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		220
	Sauc	11. 	AL AL	au	Τa	L	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		220
	SII	Ld.	пк 	a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		223
	Tha	LIA	na			•	•	•	•	•	•	•	٠	•	٠	•	•	•	٠	٠	٠	٠	٠	٠		221
	Unit	ed:	S	ta	te	S	ot	1	λшe	rı	.ca	L	•	•	•	•	•	•	•	•	٠	•	٠	•		233
ANNEX	11:	в	IB	LI	00	;R <i>A</i>	PH	Y	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		237
	Part	: 1	_	R	ec	coπ	nme	nd	led	la	ıdd	lit	ic	na	1	re	ead	lir	ıg	•	•	•	•	•		237
	Part	: 2	_	R	ef	er	en	Ce	es								•		•	•		•	•			242

EXHIBITS

3.1	Topics most commonly covered in household surveys	22
3.2	Special design requirements by topic	28
3.3	Topics suitable for a single multi-round survey	30
3.4	A design with 50 percent quarter-to-quarter and 50 percent year-to-year overlap	42
3.5	Basic design features and options for multi-subject surveys	44
4.1	Categories of information that may included in records for frame units	56
4.2	Checklist of desirable frame properties	65
4.3	Attributes used to classify frame units by urban-rural characteristics in selected countries	73
4.4	Serpentine numbering of elementary frame units within the next higher-level unit	75
4.5	Steps in planning the development of a master sampling frame	80
4.6	The population census and the master sampling frame (MSF)	82
4.7	Summary: Inventory and evaluation of potential frame inputs	90
4.8	Key characteristics of master sampling frames for selected countries	92
4.9	Steps in the initial development of an MSF	98
4.10	Frame units for some typical multi-stage designs 1	0 <u>9</u>
4.11	Steps in planning for the preparation and use of housing unit listings	16
5.1	Key features of a master sample 1	26

Exhibits (cont'd)

.

5.2	Economies of scale resulting from the use of master samples	130
5.3	Key features of master samples used in multiround surveys: Jordan, Saudi Arabia and Thailand	141
5.4	Year-to-year overlap (percent) for three master-sample designs	143
5.5	Stages of sampling for three master-sample designs	146
5.6	Sample units in order of their relative stability	156
5.7	Number of master sample units per survey or survey round for smallest publication area: selected countries	164

.

,

TABLE

4.1	Total num	ber of	administrative	subdivisions	in	Thailand,	
	by type:	1970,	1980 and 1982		•		69

CHAPTER I

INTRODUCTION

This document is one of a series of technical studies prepared for the use of countries participating in the United Nations National Household Survey Capability Programme (NHSCP). The NHSCP is designed "to help interested developing countries obtain, through household surveys and in conjunction with data from censuses and other sources, a continuing flow of integrated statistics for their development plans, policies and programmes, and in line with their own priorities. For this purpose, the NHSCP aims to assist the interested countries to develop enduring national instruments and skills for survey-taking" (United Nations, 1980b). As a country-oriented programme, the NHSCP does not propagate any fixed model of surveys. The scope and complexity of the data collection programme will differ from country to country, depending on specific needs and potentialities. However. continuity and integration of household survey activities are essential features of all NHSCP country programmes.

Previous NHSCP technical studies have covered topics such as data processing, non-sampling errors, and questionnaire design. The studies are intended to supplement the <u>Handbook of Household Surveys</u> (United Nations, 1984), which provides an overview of general survey planning and operations.

The topics covered in this technical study are:

- o The development and maintenance of sampling frames
- o Sample designs for an integrated programme of household surveys, with particular attention to the use of master samples

Both topics are of central importance in the undertaking of an integrated household survey programme (IHSP), which, along with the buildup of national capabilities and facilities for survey taking, is a primary objective of NHSCP country projects.

Many excellent texts and manuals cover the theoretical and practical aspects of sample design for a single household survey. Much less has been written about sampling and related issues that arise in designing a programme of household surveys intended to provide data on several topics over an extended time period. This is a major gap, because an effective programme of surveys cannot be designed by treating each survey as a separate undertaking. This study attempts to fill the gap by providing a convenient and practical discussion of sampling and related issues that arise in designing IHSPs for developing countries.

As discussed more fully in Chapter II, the key to a successful IHSP is <u>integration</u> -- use of the same concepts, survey personnel, facilities, sampling frames, and related materials in multiple surveys and survey rounds. Integration offers gains in efficiency and quality that cannot be realized if each new survey is designed and carried out independently of previous surveys. The NHSCP is a country-oriented programme and all countries participating are urged to develop household survey programmes whose content and design are adapted to their own specific data needs and to the resources available for household surveys. Each country, however, should aim for the development of a programme of surveys that are integrated with respect to content, facilities and design.

The development of sampling frames and master samples for use in more than one survey or survey round is one of the most important aspects of integration.

A. Audience, scope, and general approach

This study is meant primarily for statisticians in developing country statistical organizations who are responsible for the technical aspects of survey design. Survey managers and others responsible for various aspects of household surveys can also benefit from the parts of the study that deal with the broader aspects of designing a program of surveys. Survey statisticians everywhere should find the study of interest for its perspectives on the historical development and current applications of the master sample concept. The presentation 18 primarily non-mathematical; however, readers with some knowledge of elementary sampling theory and some experience with its application in the design of household surveys will find the arguments easier to follow, especially in Chapter V, which covers the use of master samples.

Illustrations and examples from ongoing survey programmes are used liberally in this study. Consideration of current practices is useful because it helps us to keep in mind the constraints imposed by developing country environments and resources. However, the inclusion of a particular case study or illustration does not necessarily mean that it represents the best possible design or procedure under the circumstances, and alternatives are suggested when considered appropriate.

The main goal of this study is to provide practical guidelines that will assist readers in planning for the development of frames and in selecting an appropriate sample design for an IHSP. It is not, of course, possible to present a detailed set of designs and procedures covering every possible set of circumstances. Readers are reminded, also, that there are many good texts and manuals that cover the design of samples for individual surveys (see Annex II). It is the intention of this study to complement rather than duplicate these materials.

In summary, the main features of this study are:

CONTEXT

- o NHSCP
- o Integrated household survey
 programmes (IHSPs)

MAJOR TOPICS	0	Frame development and maintenance
	0	Sample designs with emphasis on the use of master samples
AUDIENCE	0	Survey designers
	0	Survey managers
	0	Survey statisticians
APPROACH	o	Non-mathematical
	0	Liberal use of examples
	0	Develop practical guidelines
	0	Complement standard texts and manuals

B. Organization of this document

Chapter II introduces the key concepts and definitions used in this study. Chapter III, Designs for Integrated Household Survey Programmes, sets the stage for the discussion of sampling frames and master samples. The design of an IHSP must take into account general requirements common to all programmes, specific data requirements, and the operating environment in which surveys are to be conducted. Various combinations of data needs and operational constraints lead to different classes of IHSP designs: these are discussed in the final section of Chapter III.

Chapter IV provides a detailed discussion of sampling frames for IHSPs. The first part of the chapter covers the general nature, contents and desirable properties of frames for all stages of sampling in multistage sample designs. This is followed by a discussion of the sources of frames for IHSPs and the relative merits of using existing frames (e.g., a population census) or creating new frames. A stepby-step exposition of the process of designing a master sampling frame for an IHSP is included. The chapter concludes with a section on secondary sampling frames.

Chapter V covers the use of master samples. It begins with a brief description of the historical origin of the concept and a discussion of the advantages and limitations of using master samples. The main part of the chapter consists of a detailed discussion of the application of the master sample concept in two different contexts: (1) in a single

-3-

multiround survey and (2) in a programme consisting of several different surveys. In each case, general design issues are reviewed and examples are discussed.

Chapter VI summarizes the key issues discussed in the preceding chapters: the importance and benefits of integration in a programme of household surveys; conclusions about the desirability of using master sampling frames and master samples; and the importance of documentation and quality assurance.

In short, the topics covered in the next five chapters are:

Chapter II Basic concepts and definitions

Chapter III Designs for integrated household survey programmes

Chapter IV Sampling frames

Chapter V Master samples

Chapter VI Summary and recommendations

The study includes two annexes. Annex I presents case studies of designs used or proposed for use in several different household survey programmes, with emphasis on the sampling frames used and the use of master samples. To facilitate the comparison of different designs, a standard format is used. Most of the case-studies are from developing countries, but a few others are included to illustrate important design features. The case-studies do not all represent designs currently in use. They are all based on the publications and reports that were readily available and do not always reflect recent programme changes and design modifications. Sources of information are identified for each case-study.

The first part of Annex II is a short annotated list of publications recommended for those who require additional information on the topics covered in this study. It is followed by a full list of the references cited in the text.

CHAPTER II

BASIC CONCEPTS AND DEFINITIONS

Most readers will have some knowledge of the basic ideas of sampling and survey design and their application in household surveys. Nevertheless, in order to avoid misunderstandings, it may be desirable to review some concepts and definitions that are relevant to this technical study. The concepts and definitions to be discussed come under the following headings:

c	Integration in a system of household surveys
c	Survey units
CONCEPTS AND C	Multistage sampling
DEFINITIONS	Sampling frames
c	Master samples
c	Optimum survey design

The first topic, integration in a system of surveys, is less likely than the others to be familiar to readers, but is basic to an understanding of the objectives of this study. The next three topics -survey units, multistage sampling and sampling frames -- are all important in the design of individual household surveys. The fifth concept discussed in this chapter is the master sample. The phrase "master sample" is probably familiar to most survey statisticians, but some might have difficulty in giving or agreeing on a precise definition. It is recommended that all readers review this topic, because master samples, as defined for this study, are an important feature of many integrated household survey programs. Finally, the meaning of optimum survey design in the context of a programme of surveys (as opposed to a single ad hoc survey) is discussed.

A. Integration in a system of household surveys

First, we need to consider what is meant by a <u>survey</u> and by a system or programme of household surveys. A single household survey may be <u>ad</u> <u>hoc</u> (carried out only once) or it may be repeated on several occasions. In the latter case, it can be referred to as a periodic or continuing survey. For periodic surveys, data collection is carried out during discrete time periods, usually spaced at regular intervals. In continuing surveys, data collection is continuous. For both periodic and continuing surveys, reference is often made to <u>survey rounds</u> (or cycles). For a periodic survey, each discrete data collection period constitutes a survey round. For a continuing survey, the term survey round usually refers to the period for which separate estimates are produced. Depending on the survey data requirements and design, each round might cover from one to twelve months. A survey whose content is limited to a single subject, such as health, is called a <u>specialized survey</u>. Surveys that cover more than one major subject are called <u>multi-subject surveys</u>. Either type of survey can consist of a single round or multiple rounds.

In most periodic or continuing multi-subject surveys, at least some of the survey content remains constant from round to round. This basic or core content is often supplemented by inquiries on other topics which vary from round to round.

Not all subjects covered in household surveys can be easily combined in a single multi-subject household survey. Therefore many national statistical organizations conduct several household surveys, some of them continuous or periodic and some <u>ad hoc</u>. Thus, a national system or programme of household surveys may consist of a single multi-subject survey with multiple rounds or it may consist of two or more separate surveys. Frequently, all government household surveys are conducted by a single centralized statistical agency, but this is not always so. Whether one can think of a set of surveys conducted by two or more agencies as a system of surveys depends on the extent to which the agencies coordinate their efforts.

<u>Integration</u>, in the context of a programme of household surveys, implies linkages between surveys or between rounds of a single survey. These linkages have two objectives: to reduce the overall cost of the survey programme and to enhance the value of the survey results. They cover three aspects of the design and operation of a programme of surveys:

- o Substantive aspects. Linkages in this area include such features as: use of standard definitions of the target and survey populations (see next section); use of standard definitions and survey questions for frequently-used classifiers such as age, sex, religion, race/ethnicity, marital status, education and activity status; and coverage of multiple topics in a single survey, in the same or successive rounds, in a way that permits the data for these topics to be analyzed jointly.
- Sharing of survey personnel and facilities. The conduct of a ο household survey requires staff trained in cartography. statistical methods, field operations, data processing and subject-matter analysis. A variety of facilities such 88 transport, computers, peripheral equipment, copiers and printing equipment, are essential. It is difficult, if not impossible. to assemble competent staff and adequate facilities for a single ad hoc survey. Effective use of permanent staff and facilities in a household survey programme is a key aspect of integration.
- o The use of survey designs that make it possible to allocate some of the costs associated with frame development and sample selection over several surveys.

The main focus of this technical study will be on the last of these three aspects of integration, i.e., the development of survey designs for an integrated household survey programme (IHSP).

The concept of integration can also extend to links between household surveys and other activities carried on by national statistical organizations. There are often important linkages between household surveys and censuses of population and housing. These linkages extend to substantive aspects and sharing of facilities; however, the key linkage in the context of this study is the potential use of census materials and results in the construction of frames for household surveys. Household surveys may also be linked in various ways with economic censuses and surveys or with systems of administrative records.

In summary, the goal in designing an IHSP is to maximize desirable linkages:

Within the household survey programme

0

- Shared concepts, definitions and content
- Shared personnel and facilities
- Shared costs of sample selection
- Between population and housing censuses and the household survey programme

The objectives of integration, thus defined, are to reduce the costs of surveys and to enhance the utility and quality of results.

B. Survey Units

To design a survey, one must first define one or more <u>target</u> <u>populations</u>. The definition of a target population consists of two parts: the kinds of units to be covered and the extent or limits of coverage of those kinds of units.

The kinds of units most commonly covered in household surveys are persons, households, families and housing or dwelling units. In addition, target populations for household surveys sometimes include economic enterprises associated with households, such as agricultural holdings or household industries.

For each type of unit, there are usually some specific questions about the extent or limits of intended coverage. Some units may be excluded from the target population because the topic of the survey does not apply to them. Thus, a labour-force survey might exclude children under 14 (or some other age) and a housing survey might exclude some kinds of institutional living quarters.

Some units may be excluded from a survey because they are costly or difficult to cover. Examples are citizens living abroad and members of

unassimilated populations living in remote areas. The term <u>survey</u> <u>population</u> is sometimes used to describe the part of the target population remaining after making these exclusions for practical reasons.

The units to be included in the target or survey population are called <u>elementary units</u>. In a household survey, data on the characteristics and behavior of a sample of these units are collected and used to make estimates for the survey population. The goal of the sample design for the survey is to produce valid estimates, with measurable reliability, for that population.

Many surveys are designed to cover more than one survey population. A single household survey might, for example, yield estimates for persons, households and housing units.

Sampling units may or may not be the same as the elementary units. Sampling for household surveys usually proceeds in more than one stage (see section on multistage sampling later in this chapter). At the first stage of sampling the units from which the sample is selected are usually areas with defined boundaries. They may be political subdivisions of a country or specially-defined areas such as the enumeration areas used in a census. Area units may also be used as sampling units at the second and later stages of sampling and sometimes even at the final stage. In other surveys, however, the final stage of sampling may be the selection of households or housing units from listings prepared for a sample of area units selected in the previous stage.

Because the sampling units are not necessarily the same as the elementary units in the survey population, one must often establish <u>rules</u> <u>of association</u> (sometimes called counting rules) to link the two kinds of units and to ensure that each elementary unit in the survey population has a known (or knowable) probability, not zero, of being included in the sample. For example, if the sampling units selected at the final stage of sampling were small area segments, a simple rule of association would be to include in the sample all housing units located in the selected segments. A somewhat more complex rule might be required to associate households with a sample of housing units selected from a listing. The rule would have to take into account the possibility that some housing units may be occupied by more than one household and, conversely, that some households may occupy more than one housing unit.

Most rules of association are unique in the sense that every elementary unit is associated with one and only one sampling unit. However, it is possible to develop valid probability sample designs using rules that associate some or all of the elementary units with more than one sampling unit. This technique, which is called network sampling or sampling with multiplicity, can be especially useful in sampling rare populations, such as persons with a specific disease (for further details see Sirken, 1970, 1975). To summarize this discussion of units:

- o Survey target populations are made up of <u>elementary units</u> such as persons and households.
- o The sample selected for a survey consists of <u>sampling units</u> which may be either areas or units such as housing units or households selected from a listing.
- o If the elementary units and the sampling units are not the same, rules of association must be developed to link them.

C. Multistage sampling

If costs were not a factor, the ideal sample design for a household survey might be to prepare a current listing of housing units or households for the entire country and to select random or systematic samples from the listing. This is not practical for two reasons: first, the cost of preparing such a list would be exorbitant and second, much of the interviewer's time would have to be spent travelling from one sample unit to another between interviews.

Such cost considerations lead to the use of multistage sampling procedures in which sampling is carried out in two or more stages. A simple example of a two-stage design would be:

- Stage 1. Select a sample of enumeration areas (EAs) as defined for the most recent population census.
- Stage 2. Prepare housing unit (or household) listings for each sample EA (this might be done by updating census listings if they are available) and select a sample of housing units (or households) from each one.

In a three-stage design, a sample of administrative districts, each containing several EAs, might be selected at the first stage. The second and third stages of sampling would be the same as stages 1 and 2 in the previous example, except that the selection of EAs in what is now stage two would be made only in the sample districts selected in the first stage.

A key feature of multistage sampling is that the sampling at each stage after the first is restricted to the sampling units actually selected in the previous stage. This reduces the resources needed to prepare sampling frames (see next section) for the second and succeeding stages of selection. Another important feature is that the sample units to be surveyed will be clustered, rather than widely dispersed throughout the entire area occupied by the survey population. Clustering of sample units usually increases the level of sampling errors for a sample of fixed size. However, savings resulting from lower frame development and interviewer travel costs often permit the use of samples large enough to more than offset the effects of clustering on sampling errors.

The sampling units used at the first stage of sampling are called <u>primary sampling units</u> (PSUs). Those used at the final (ultimate) stage are called <u>ultimate sampling units</u> (USUs). In designs with three or more stages, units used at the intermediate stages are called secondary (or second-stage) sampling units (SSUs), tertiary (or third-stage) sampling units, and so on. Thus, in the three-stage example given above, the sampling units are:

PSUs: Districts SSUs: Census EAs USUs: Housing units (or households)

D. Sampling frames

A <u>sampling frame</u> is a listing (explicit or implicit) of the units from which the sample selection is to be made at any stage of sampling. As pointed out earlier, the units in the frame may be either areas or units such as households or housing units. In this connection the terms area frame and list frame are often used.

In multistage sampling, a frame is needed for each stage of sampling. In the two-stage design of the previous section, a list of census EAs would be needed for the first stage of sample selection. Lists of housing units or households would be needed for the second stage, but only for the sample EAs. In this study the term <u>secondary</u> <u>sampling frame</u> will be used for frames that are developed specifically for the second and subsequent stages of sample selection.

Any sampling frame used for the first stage of selection must cover the entire survey population. Frequently such a frame will be used to select samples for several different surveys or for use in different rounds of a continuing or periodic survey. Such frames are referred to in this study as <u>master sampling frames</u>. The careful construction and maintenance of a master sampling frame are important elements of an IHSP.

Frames are discussed in detail in Chapter IV.

E. Master Sample

Given the existence of a master sampling frame, it would be possible to select the samples needed for different surveys or survey rounds entirely independently of each other. However, there are important potential benefits from performing the initial stages of selection in such a way that the resulting samples can serve the needs of more than one survey or survey round.

For this study, a <u>master sample</u> is defined as a sample from which subsamples can be selected to serve the needs of more than one survey or survey round. A master sample can take several forms. It may consist of a sample of PSUs from which subsamples are selected as needed. Various options are available for selecting subsamples needed for individual surveys or survey rounds. The selection of subsamples may be: entirely independent, designed to avoid any overlap or designed to produce a specified proportion of overlap. A master sample could also consist of SSUs or USUs, with the selection of subsamples being restricted to the lowest stage units included in the master sample.

A special type of master sample consists of two or more samples of units selected independently in one or more stages, using the same design. These independently selected samples are called <u>replicates</u>. In this case, the first stage of subsampling from the master sample would consist of the selection of one or more replicates for each survey or survey round, with or without overlap, as desired. Suppose, for example, that a master sample consisted of 40 replicates, each being a one-stage sample of 100 census EAs. For some purposes, an appropriate design for a multi-round survey would be to use replicates 1 to 4 in round 1, replicates 2 to 5 in round 2, replicates 3 to 6 in round 3, and so on.

The design and uses of master samples are discussed in detail in Chapter V.

F. Optimum survey design

The number of ways in which one can design a household survey to meet a specific set of data requirements is virtually unlimited. Historically, the idea of optimum design started with the sample design for a single survey and dealt with such design features as choice of sampling units, stratification, assignment of selection probabilities, selection method and estimation procedure. The objective was to develop a sample design that would meet reliability requirements at the lowest possible cost, or alternatively, to produce the most reliable estimates for a fixed expenditure of resources. Later, the concept of total survey design came to the fore. Survey practitioners became more aware of the impact of non-sampling errors on the quality of survey results. In addition to deciding on sampling procedures, survey designers had to make choices among alternative data collection modes, respondent rules, callback procedures, wording and format of questions, computer edit rules and a variety of quality assurance techniques at all stages of the survey. The possible use of higher-cost procedures designed to reduce non-sampling errors had to be weighed against reductions in sample size that might be needed to cover the higher costs. The preferred criterion for comparing alternate designs shifted from the sampling errors to the mean square errors of survey estimates, at least in those situations for which quantitative estimates of the latter could be obtained (Fellegi and Sunter, 1974).

Most of the literature on survey design, both theoretical and applied, addresses the question of how to optimize the design of a single survey. This is true of most of the texts on survey sampling, although there are some exceptions. The first edition of Yates (1949) has a brief section on master samples and also discusses estimation procedures for sampling on two or more successive occasions with partial sample overlap. The latter question is also addressed by Hansen, Hurwitz and Madow (1953). Kish (1965) has a more extensive discussion of issues arising in multiple surveys; sections of his text cover topics such as repeated selections from a listing, correlations from overlaps in repeated surveys, panel studies and designs for measuring change, continuing sampling operations and changing selection probabilities. A11 of these topics are important in the design of an IHSP.

The point that does not seem to have been made explicit until more recently, is the importance, to a national statistical organization, of planning a programme of household surveys, as opposed to ad hoc design of The benefits of integration, which were described individual surveys. earlier in this chapter, cannot be fully realized unless there is a deliberate effort to design an IHSP, covering a period of several years. For example, the development of a high quality sampling frame is expensive and the costs could not be justified if the frame were to be used in only one survey. The cost of developing and maintaining a master sampling frame for use in a continuing program of surveys is, however, much easier to justify, since it can be spread over several surveys. Similarly, the use of a master sample may make it possible to decrease the costs of sample selection, including the preparation of frames for the second and subsequent stages of selection, attributable to each Thus, the principles of optimum or total survey design, when survey. applied in the context of a programme of surveys, can lead to designs that differ substantially from those that might be developed for independently-designed surveys.

Designing an IHSP is a complex undertaking and success cannot be guaranteed by following a few simple rules. Each country's data requirements and operating environments are different and must be taken into account. Nevertheless, the potential benefits of integration are great: this is why integration of surveys is strongly emphasized in the NHSCP. The aim of this technical study is to present some ideas and examples that will contribute to the development of optimum designs for IHSPs.

CHAPTER III

DESIGNS FOR INTEGRATED HOUSEHOLD SURVEY PROGRAMMES

This chapter examines factors that must be considered in developing a plan for an integrated household survey programme (IHSP). These factors fall into three major areas:

PLANNING	o General design requirements
CONSIDERATIONS	o Data requirements
	o Operating environment and constraints

These topics are discussed, in the order shown, in the first three sections of this chapter.

The IHSP designs that have been developed by different countries show wide variation in the extent of integration of the collection of data on different topics and in the methods of integration. To a considerable degree, this variation results from real differences in the data requirements and operating environments of different countries. To some extent, also, these design differences reflect the predilections of the individuals who designed the survey programmes. It should not be thought that there is only one design for an IHSP that can meet a country's needs effectively. The particular design selected will work if it is carefully thought out and if the statistical organization is well-managed and committed to making the programme work.

The final section of this chapter identifies three general classes of IHSP designs which are frequently used. As will be seen from the examination of case-studies, not all IHSP designs fit neatly into one of these three categories; some of them use features taken from two or even all three approaches.

Chapters IV and V cover detailed aspects of the broad classes of designs introduced in the final section of this chapter. Chapter IV is about frames, with emphasis on master sampling frames, and Chapter V describes the application of the master sample concept in IHSP designs.

A. General design requirements

General design requirements are those that should apply to all IHSP designs, regardless of the particular kind of design selected. They include:

- o Long-range planning
- o Coordination with the population census
- o Flexibility
- o Use of probability sampling
- o Documentation

Each of these requirements is discussed in this section.

1. Long-range planning

The design of an IHSP consisting of a multi-round survey or multiple surveys requires long-range planning. The development of a plan for surveys and survey operations covering a period of several years is the only way to realize the benefits of integration and to ensure that the necessary personnel and facilities will be available when needed. A broad survey plan is a prerequisite to working out a detailed sample design.

How many years should a plan cover? Typically, initial country plans for NHSCP projects specify the surveys to be conducted and the topics to be covered for a five-year period. These plans are usually quite detailed and firm for the first year or two. Beyond that point they tend to be more tentative because of uncertainties about future data requirements and availability of resources. Planning is a continuous process. As time passes, plans should be updated to cover additional years and to firm up the details for surveys that are about to begin. Past experience should be reviewed to ensure that future plans are realistic.

As the time approaches for a census of population, the household survey plan should provide for close coordination of census and survey activities. This issue is discussed in subsection A, 2, below.

Broadly speaking, the survey plan should cover: the timing and data requirements for each survey and survey round; the staffing requirements, especially for field work and data processing; the facilities needed, such as transport and data processing equipment; the sampling operations to be performed, such as field listing and the development and updating of a master sampling frame; and, at least for the immediate future, the scheduling of survey operations, such as pretesting, listing, data collection, manual reviews and edits, data entry, computer edits, tabulations, and publications. All of these features interact in the development of the overall plan and the sample design. Perhaps the logical place to start is with the data requirements, but as the other elements of the plan and the IHSP design are worked out it may turn out to be necessary to revise the data requirements. The question of how to meet professional staffing requirements deserves special attention in the planning stage. Some skills needed to conduct surveys may be lacking or in short supply: for example, many statistical organizations do not have enough persons trained in sampling, data processing and analysis of survey data. In the short run these gaps can be and frequently are filled by bringing in outside advisers as needed. Requirements for particular kinds of technical assistance must be scheduled and made known to potential donor agencies or other sources well in advance to ensure the availability of qualified advisers when they are needed. In the longer run, of course, each country's aim should be to arrange for its survey staff to receive the training and experience needed to become self-sufficient in all aspects of household survey operations.

2. Coordination with the population census

Most countries participating in the NHSCP conduct population censuses at regular intervals, generally every ten years. As mentioned in section A of the preceding chapter, integration, as it applies in an IHSP, should include both internal linkages between surveys and external linkages with the population census. These linkages affect three major areas: content, shared resources and frames.

With regard to content, there are usually several variables that are common to the census and some or all of the household surveys. Typically, items like age, sex, family relationship, marital status, educational attainment and size of household are included in the census and are also used in most surveys as classifiers. Other census items, like activity status, occupation and industry, are likely to be included in some of the household surveys. Definitions of urban and rural areas and geographic regions are usually needed for both censuses and surveys. It is very desirable, from the data users' point of view, that these variables be defined in the same way or at least in a compatible manner for the census and for all surveys in which they appear. Not only should the definitions be compatible, but whenever possible the same question wordings and response formats should be used.

Such standardization of variables has significant benefits:

- o It facilitates analyses requiring the use of data from different surveys or from a census and one or more household surveys.
- For the surveys, it makes possible the use of efficient ratiotype estimators based on census totals for PSUs or population subgroups.
- It facilitates evaluation of comparative coverage in censuses and surveys, for example, by comparing survey estimates of population by age and sex with independently-derived projections of census results.

With regard to shared resources, it is almost certain that some of the same staff and facilities will be used for the population census and the household surveys. The survey plan needs to take account of the special needs of the census. In all likelihood, survey data collection will be suspended or its level considerably reduced during the census enumeration period.

For this technical study, the most important linkage between population censuses and household surveys pertains to the frames used for both. The use of the population census frame along with census results to produce a master sampling frame for household surveys will be discussed in detail in Chapter IV. One point that will be emphasized is the need, in planning for a census of population, to bear in mind that one of the products of the census should be the raw materials for a master sampling frame for household surveys. If no thought is given to this requirement until after the census enumeration has been completed, the census materials available for the development of the frame will probably prove to be less than ideally suited for that purpose.

3. Flexibility

Flexibility in IHSP designs is desirable because it is impossible to anticipate all of the requirements for data from household surveys that may arise in the course of the period (say five years) covered by the programme plan. Unanticipated needs occur for reasons such as changes in national priorities and unpredictable events affecting specific population groups or sectors of the economy.

Flexibility means the ability to respond quickly to unanticipated data requirements and to do so in a way that does not delay or otherwise interfere with ongoing survey operations. There are several features that can be included in IHSP designs to provide increased flexibility. One possibility is to leave some "open space" for new topics in selected rounds of a continuing or periodic multi-subject survey. In India, for example, a ten-year plan for topics to be covered in the National Sample Survey designated specific topics for seven of the ten years. The other three years were left open for topics of current interest (Rao and Sastry, 1975).

The availability of a well-designed and maintained master sampling frame, especially one that is computerized, adds to flexibility, since the frame can be used quickly to select new samples or to supplement existing samples as needed. Even more desirable would be the availability of a master sample with reserve units that can be used for samples needed to meet unexpected requirements, especially if the reserve units are USUs such as housing units, households or compact area segments. A master sample can be deliberately designed with the capacity to provide samples for surveys other than those planned at the time the master sample is selected. For example, a general-purpose sample design proposed for Jordan called for the selection of a master sample consisting of 21 replicates, even though only a subset of these replicates would be needed for the first round of the Multi-Purpose Household Survey and a specific plan for sample rotation in subsequent rounds had not yet been developed (Jordan case-study).

There are, of course, costs associated with the selection of samples to be held in reserve to meet unexpected needs. The costs of selecting additional samples of PSUs or SSUs that are already identified in a master sampling frame (as was proposed in Jordan) are relatively small. However, if selection proceeds to the stage where field-work is needed to subdivide areas or to list housing units, significant costs can be incurred. Furthermore, listings can become outdated fairly quickly in some areas so that after some time has passed they cannot be used without updating. The problem then, is to find ways of building flexibility into the IHSP design without using resources unproductively. Methods of doing this are discussed in Chapter V.

4. Use of probability sampling

Probability sample designs are used in virtually all household surveys conducted by national statistical organizations. The basic principle of probability sampling is that each member of the target population should have a known probability, not zero, of being selected. This requires the use of random selection procedures at all stages of sampling.

Unless probability sampling is used, it is not possible to estimate sampling errors directly from the sample data. Probability sampling avoids biases that may be introduced unintentionally if purposive or judgment samples are selected. It also lends credibility to the survey results: the statistical organization that uses carefully documented probability selection procedures cannot legitimately be accused of tampering with the sample selection process in order to manipulate survey findings. The use of probability sampling is strongly recommended for NHSCP participants.

Sometimes certain areas or population groups are deliberately excluded from the sampling frame for a household survey on the grounds that data collection is too hazardous or has unacceptably high costs. Such exclusions do not violate the principles of probability sampling; they should, of course, be clearly explained to users of the survey results. On the other hand, making substitutions for sampled units for which it proves impossible to collect information is a violation of these principles and is not recommended (for a fuller discussion, see United Nations, 1982a, pp 95-96).

5. Documentation

The importance of full and accurate documentation of procedures and outcomes at all stages of survey work is difficult to overstate. It is critical in connection with the development and maintenance of a master sampling frame and the design, selection and use of master samples. To give an idea of the kinds of documentation needed, following are some examples:

For the development and maintenance of a master sampling frame

- o A standard record format should be developed for each type of unit (district, village, enumeration area or block, etc.) included in the frame, with identifiers, data items, such as census population or household counts, and other information, such as map keys. A record layout should be prepared, along with detailed source information for each field of the record.
- o A reliable system should be established for learning about the creation of new political subdivisions and changes in existing ones. Procedures should be developed for correcting frame records to reflect such changes and an accurate record should be made of each correction. It may also be necessary to split, combine or otherwise adjust frame units because of substantial changes in population. All such actions affecting frame units must be carefully documented so that any necessary adjustments can be made in previously selected samples. Maps associated with the master sampling frame must be carefully annotated or revised to reflect such changes.

For the selection and use of a master sample

- o The procedures used to select sample units at each stage of selection should be fully described in writing. The sampling worksheets or computer listings actually used in the selection should be carefully preserved. As a safeguard, one or more additional sets of these worksheets should be made.
- o An accurate record showing which master sample units have been used in samples for particular surveys is essential. Such a record makes it possible both to establish full or partial sample overlap between surveys or survey rounds when desired and also to avoid placing undue burden on particular respondent groups by including them in too many samples.

In designing documentation for sampling activities, two considerations are paramount. First, a standard identification numbering system for frame units and sampling units is essential. Every record associated with a particular unit should include an identifier that will allow it to be linked readily with any other records for the same unit. The numbering system for frame units should be designed to accommodate changes in these units.

Second, the documentation system must provide the information needed to determine the overall selection probability of each USU included in the sample for every survey. Only in this way can the correct weights be applied to each unit in the sample. If part of the selection is done in the field, a special effort will be necessary to ensure that each person involved submits full and accurate information for the selection operations performed.

Good documentation is part of the investment required to realize the benefits of integrated household survey activities. Without it, there is a real danger that some of these benefits will be lost.

B. Data requirements

A logical first step in developing a plan for an IHSP is to determine, at least tentatively, the subjects to be covered and, for each subject, the frequency with which data are needed and the specific kinds of data desired. It is also useful to assign relative priorities to the topics selected.

Once a proposed set of topics has been developed, each topic should be analyzed to identify special survey design requirements associated with it. Does it require use of field staff with special qualifications or equipment? Is there significant seasonal variation to be taken into account in scheduling the data collection? Can the data be collected in a single visit to each household, or will multiple visits be required? Will an especially large sample be needed to provide subnational estimates or because the topic applies only to a small proportion of households?

The third and final step in the analysis of data requirements is to examine the topics chosen to determine which ones might be grouped in specific surveys or survey rounds. Two questions are relevant. First, are there topics which should be investigated for the same sample of households so that data on these topics can be linked for analytical purposes? Second, are there topics whose design requirements are sufficiently compatible that they might be covered in the same survey or at least in different rounds of a continuing multi-subject survey?

In summary, there are three main steps in analyzing the data requirements for an IHSP:

	0	Make a tentative selection of topics
ANALYSIS OF DATA REQUIREMENTS	0	Identify special design requirements for each topic

o Form groups of compatible topics

Each of these steps is discussed below. For additional details, readers may want to consult the <u>Handbook of Household Surveys</u> (United Nations, 1984) and a paper presented by the United Nations Statistical Office at the 1983 session of the International Statistical Institute (United Nations Statistical Office, 1983).

The choice and grouping of topics resulting from this three-step analysis should still be regarded as tentative. Analysis of the operating environment and constraints (section C, below) may lead to reconsideration of the choice of topics and their frequency and depth of coverage.

1. Topics for household surveys

The number of topics that can be investigated in household surveys is limited only by the imagination of survey planners. However, the experience of recent decades shows that the topics covered most frequently by household surveys in developing countries fall into certain fairly well-defined subject groups.

Exhibit 3.1 lists commonly-surveyed topics in a format designed to facilitate analysis of the relationships between topics. The topics are identified in broad terms. For each topic selected, choices are required concerning specific subtopics to be covered, and it will be found that these subtopics do not always have the same survey design requirements associated with them.

The basic demographic and social items (category A in Exhibit 3.1) include characteristics like age, sex, racial or ethnic group, marital status, literacy and educational attainment. These and other similar items are collected routinely in nearly all household surveys. They are used as classifiers or explanatory variables in presenting and analyzing data on most topics and to facilitate linkages between data from different surveys and between census and survey data. They may also be used as a component of the survey estimation procedure. In fact, they do not really constitute a topic for a particular survey, but are shown here for completeness.

A review of the topics expected to be covered in IHSP's by 17 countries participating in the NHSCP (United Nations Statistical Office, 1983) showed that all of the countries included labour force, income and expenditure in their programmes. All but one country included one or both of the two demographic topics (categories B,1 and B,2 in Exhibit 3.1).

Many topics are covered less than annually by most countries. Data associated with some topics, e.g., housing characteristics, are relatively stable over time; therefore, coverage more than once every three to five years may be unnecessary. Substantial resources are needed to collect and process data of good quality on household income and

EXHIBIT 3.1 Topics most commonly covered in household surveys

- A. Basic demographic and social items
- B. Demographic and social topics
 - 1. Components of population change: births, deaths, migration
 - 2. Other demographic: e.g., fertility, family planning
 - 3. Health and nutrition: e.g., current health and nutritional status, availability and use of facilities, food consumption
 - 4. Housing characteristics
 - 5. Status and activities of special population groups: e.g., youth, women, aged
- C. Socio-economic topics
 - 1. Labour force: employment, unemployment, underemployment
 - 2. Income and expenditures
 - 3. Household enterprises
 - a. Nonagricultural
 - b. Agricultural

expenditures, so that it may be beyond the capability of the national statistical office to provide annual data, even though users might want them.

Many countries collect labour force data annually and some collect it more frequently, e.g., semi-annually or quarterly. Countries that rely on household surveys as a primary source of data on agricultural production, e.g., Ethiopia, normally collect such information every year.

After drawing up a tentative list of topics and subtopics and deciding how frequently each one should be surveyed, priorities should be established, taking into account the strength of various needs expressed by potential users of household survey data. It may not be possible, at the outset of an IHSP, to cover all of the selected topics on a regular schedule. Also, compromises or tradeoffs may be required in designing surveys to cover multiple subjects; such compromises should favour the topics with higher priorities.

2. Special design requirements

The choice of appropriate survey designs is strongly influenced by the nature of the topics and subtopics to be investigated. Some topics require specially qualified field staff or special equipment; some require multiple interviews of sample households at carefully-timed intervals; some require that interviews be conducted at certain times of the year. Some topics and their associated data needs impose special requirements on the sample size and design.

A well-trained and experienced field staff can deal adequately with most household survey topics, given a reasonable amount of instruction on the subject matter. Some topics are more complex than others and require more extensive training. Income and expenditure surveys belong to this category, as do agricultural surveys that require the use of objective measurement techniques to estimate crop areas and yields. Special training would be necessary for some kinds of health and nutrition surveys, e.g., those requiring actual weighing of food consumed or physical examination of household members.

There are a few topics that cannot always be adequately dealt with by the regular field staff even with additional training. For surveys covering contraceptive practices, some countries require that all of the interviewers be women. Some health subtopics may require the use of physicians or other medically-trained personnel as interviewers.

Topics or subtopics that may require special equipment include food consumption, nutritional status and agricultural production. The cost of the equipment needed for objective measurement can affect the sample size and distribution. The timing of surveys and survey rounds depends to a considerable degree on the nature of the topics to be covered. Seasonal variations in labour requirements dictate that labour force surveys which use short reference periods be conducted either continuously during the year or in two or more rounds spread through the year. Surveys of agricultural production, especially those using objective measurement methods, must be carefully timed in relation to planting and harvesting seasons. Note also that these seasons may differ substantially in different regions of a country. Other topics and subtopics for which seasonal variation may need to be taken into account include: income and expenditures, some kinds of household enterprises, school attendance, food consumption and health status.

For some topics, the collection of sufficiently accurate data may require multiple visits to the same sample households. Surveys of expenditures, food consumption and income in particular, often involve multiple visits. Decisions on whether to interview the same households more than once depend on careful consideration of the lengths of reference periods required for analytical purposes and the ability of respondents to recall particular kinds of events or transactions and to place them accurately in time. Multiple visits place a greater burden on sample households; this factor should also be considered.

Sample size requirements are most directly affected by whether or not separate estimates are wanted for political divisions of a country, such as regions, provinces, or states. For large countries with significant regional variations in climate, ethnicity and economic activities, subnational data are of considerable importance for most if not all of the major household survey topics and are likely to be part of the data requirements for an IHSP to the extent that resources permit.

The importance of economic topics varies between urban and rural areas. Although a few agricultural households may be in areas classified as urban, most of them will be concentrated in rural areas and in the smaller urban places. The kinds of labour force information needed may differ for urban and rural areas. For the former, it may be important to track trends in unemployment with annual or more frequent estimates; for the latter, seasonal underemployment may be the main concern, and an observation once every three to five years might be adequate for policymaking. Income and expenditures are usually of interest for all areas, but the variation between households in rural areas is likely to be considerably smaller than in urban areas, so that the desired level of reliability can be achieved with smaller samples in rural areas.

Surveys aimed at rare items or infrequent occurrences, i.e., those that are only found in a fairly small proportion of households, require large samples, at least for the purpose of "screening" households to identify those that are part of the target population. Examples of such topics might be disability or vocational education and, to a lesser extent, births and deaths. There are various ways to obtain adequate samples of rare populations. If a country uses a sample design with listing and subsampling at the final stage of selection, screening for the rare items can be done as part of the listing operation. Another technique is to accumulate data over two or more rounds of a continuing or periodic survey. If lists covering some members of the target population can be obtained from some type of administrative register, multiple-frame sampling is a possibility.

From this review of special design requirements, it may be evident that household surveys of the agricultural sector have a number of requirements that set them apart from most other topics. They can place heavy seasonal demands on the field staff. The allocation of sample households best-suited to agricultural surveys may differ substantially from that which is optimum for most other topics. It may be necessary to supplement the household survey agriculture data with separate coverage of units such as estates, plantations or other types of agricultural enterprises in the non-household sector.

Nevertheless, there are several countries for which household agricultural activities are a major component of the total economy, so that knowledge of their structure, inputs and outputs is essential and receives a high priority in planning an IHSP. Several African countries (Ethiopia, Kenya, Lesotho, Malawi, Mali, Zambia, and Zimbabwe) have decided to make annual household surveys of agriculture the core of their NHSCP projects (United Nations Statistical Office, 1983).

To summarize, an important step in the analysis of data requirements is to identify the special design requirements associated with each topic or subtopic chosen for potential inclusion in an IHSP. Requirements to look for are:

- Need for special training of interviewers
- Need for interviewers with special qualifications
- o Need for equipment for objective measurement

SPECIAL O DESIGN REQUIREMENTS

- o Seasonality requiring special timing of surveys
 - o Need for multiple visits to sample households
 - o Need for subnational estimates
 - Uneven geographic distribution of target population
 - o Rare items or infrequent occurrences

3. Grouping topics with similar requirements

The final step in the analysis of data requirements preparatory to the development of survey designs for an IHSP is to examine the datalinkage requirements and the special design requirements for all of the topics and subtopics chosen and to group topics and subtopics in appropriate ways.

As mentioned previously, the basic demographic and socio-economic items play an important role in the analysis of almost every topic covered in household surveys; consequently, these topics are included in virtually all surveys.

The major economic topics have significant linkages among themselves for data analysis. The employment of family members provides a major input to household enterprises and the outputs of these enterprises are a major determinant of household income. Consequently, in countries or areas where a high proportion of the population is employed in household enterprises it may be desirable to cover these topics in the same survey.

The demographic and social topics are perhaps less closely interrelated; nevertheless, some kinds of analysis would call for the inclusion of different topics in the same survey. For example, it might be desirable to investigate relationships between housing characteristics, such as sanitary facilities and sources of drinking water, and health status.

There are also, of course, linkages between socio-economic and social or demographic variables. In particular, household income is often used as a classifier or explanatory variable in analyses of fertility, mortality, health status and other demographic and social variables.

Direct linkages between different topics are valuable to users, but there are some limits to the number of topics and subtopics that can be covered in a single survey. The patience and goodwill of respondents should not be abused by conducting excessively long interviews. It is also important that the length and complexity of survey questionnaires not exceed what can be handled by the data processing staff on a timely basis.

These limitations suggest the need to establish priorities for the topics and subtopics to be linked. When topics are linked, some of them may be investigated in less detail than if they were being covered in separate surveys. Inquiries about income, for example, may be quite detailed in a survey devoted entirely to income and expenditures. On the other hand, if income is needed in connection with analysis of demographic and social variables, a less detailed inquiry may be sufficient to provide a rough classification of households by income. Linkages of topics and subtopics for analytical purposes must be selected on the basis of user requirements. Many such linkages are possible; among those most commonly made are:

Basic demographic and social items $\leftarrow -\rightarrow$ All topics

Labour force $\leftarrow - \rightarrow$ Income $\leftarrow - \rightarrow$ Household enterprises

Income(short version) $\leftarrow \rightarrow$ All demographic and social topics

It is also desirable to group topics or subtopics that have similar design requirements. Potential groupings can be identified by constructing a two-way table or matrix, showing the selected topics or subtopics in the heading of the table and types of special design requirements in the stub. Using such a table, a column-by-column comparison will make it relatively easy to spot groups of topics with similar requirements.

Exhibit 3.2 illustrates a format suitable for this kind of analysis. It shows the special design requirements associated with the major topics commonly investigated in household surveys. In practice, only those topics chosen for coverage in an IHSP would be included in the analysis. The cell entries in Exhibit 3.2 are based on subjective evaluation of past experience. For example, all topics require a reasonable amount of interviewer training, however, certain ones are identified as requiring special training: health examination, food consumption and crop areas and yields because they involve the use of special objective measurement techniques; and income and expenditures because of the diversity and complexity of the items to be covered.

A review of Exhibit 3.2 makes it evident that the topics and subtopics most affected by special design requirements are:

Topic Health and nutrition	<u>Subtopics</u> Health status as determined by examination
	Food consumption (when objective measurements are used)
Income and expenditures	A11
Household enterprises	Crop areas and yields (when objective measurements are used)

This analysis makes it clear why separate surveys are usually conducted for these topics and why, frequently, special groups of interviewers are selected or trained for such surveys. Surveys on fertility and family panning are often conducted separately because of the need to use only female interviewers. As a rule, surveys on topics
		Demographic and t	social topics			Soc	io-economic topic	8
Special design requirements	Populstion change	Other demographic	Health and nutrition	Housing	Special population groups	Labour force	Income and expenditures	Household enterprises
Special interviewer training requirements	N	N	Yes, for health examination, food con- sumption	ON	ON	ON	Yes, due to complexity of subject matter	Yes, for crop areas and yields
Special interviewer qualifications	O N	Yes, for fertility and femily planning	Yes, for health examination	NO	ОИ	ON	ON	ON
Special equipment for objective measurement	O N	ON	Yes, for health examination, food consumption	OH	ON	N O	ON	Yes, for crop areas and yields
Seasonality may requir special timing of surv	эув NO	NO	Yee, for health Btatus and food consumption	0 x	Yes, for items on school attendance	YES	YES	YES
Need for multiple visi to sample households	ta YES	NO	Yes, for food consumption	ON	ON	ON	YES	Yes, for crop areas and yields
Uneven distribution of target population	0 N	Ю	NO	М О	NO	ON	МО	Yes, especially for agriculture

Erhibit 3.2 Special design requirements by topic

involving special interviewer training or qualifications, use of special equipment or multiple visits use smaller samples because of the costs of meeting these special design requirements.

Looking at the other side of the coin, what topics or subtopics have design requirements sufficiently alike that they can readily be covered in a single survey or in different rounds of a continuing or periodic multiround survey with a more or less fixed sample design? Exhibit 3.3 provides a list of topics which might reasonably be included in a single multiround, multi-subject survey.

The topics covered in such a survey do not necessarily need to be restricted to those shown in Exhibit 3.3. As suggested earlier, inclusion of a simple inquiry on household income might be included in some rounds because of its value for analytical purposes. In a country where agricultural or nonagricultural household enterprises are fairly common, they could be covered in one or more survey rounds. For agricultural enterprises, the inquiries would probably cover all aspects except the determination of crop areas and yields by objective Some other topics not specifically listed in measurement techniques. Exhibit 3.1 would also be suitable for this type of multiround, multi-subject survey, e.g., possession of appliances, household energy use, leisure time activities and access to and use of communications media (television, radio, newspapers, etc.).

After the selected topics have been arranged in tentative groupings on the basis of analytical linkage requirements and similar special design requirements, the next step in the design of an IHSP is to appraise the resources available for surveys and the environmental constraints that can influence the scope and design of the surveys that are being considered for the programme. This appraisal is discussed in the next section.

C. Operating environment and constraints

The analysis of data requirements discussed in the previous section of this chapter leads to the creation of a "wish list" describing the kinds of household survey data that the national statistical office would like to obtain as products of an IHSP over a period of several years. The analysis will have identified sets of topics that are compatible with respect to sample design requirements and priorities will have been assigned, hopefully on the basis of consultations with data users, to different sets of topics.

To decide how far wishes can be transformed into reality, IHSP planners must undertake a realistic analysis of the operating environment in which household surveys will be conducted and of the constraints, or limits, on the resources that are currently available or are likely to be

EXHIBIT 3.3 Topics suitable for a single multi-round survey

Topic	Sub-topic	Remarks
Basic demographic and social items		Include in all rounds
Population change	Births, deaths, migration	Births and deaths may require multiple visits
Other demographic	Fertility	Excluding sensi- tive items such as contraceptive practices
Health and Nutrition	Use of facili- ties, health status	Simple health status items not requiring spe- cial training or equipment
Housing	A11	
Special population groups	A11	
Labour force	A11	

•

available for household surveys. This analysis will influence the IHSP design in two ways. First, it will lead to a realistic view of the scope of the proposed IHSP. How many separate surveys and survey rounds can be conducted? How many different topics can be surveyed during the period covered by the plan? Will it be possible to use samples large enough to provide reliable subnational estimates?

Second, the review of environmental features and resource constraints will influence choices among alternatives for key features of the survey design. such as the number of stages of sampling, choice of sampling units at each stage and sizes of ultimate clusters in different strata.

The elements of the analysis described in this section can be grouped into seven broad categories, the first three covering different aspects of the operating environment and the next four relating to resources available for household surveys. The seven categories are:

OPERATING ENVIRONMENT 1. Administrative structure of the country 2. Target population characteristics 3. Access to sample units RESOURCE AVAILABILITY 4. Sampling materials 5. Field staff 6. Data processing capabilities 7. Technical and managerial staff

The two key questions throughout this discussion of characteristics of the survey operating environment and resource constraints are:

- How will they affect the scope of the programme? 0
- How will they influence decisions about specific design features? 0

1. Administrative structure of the country

Most countries have several kinds of political divisions and subdivisions, established to carry on the functions of government at various levels. At the first level, countries are frequently divided into relatively large units, such as provinces or states. These units may in turn be divided into smaller administrative units. Sometimes administrative divisions and subdivisions are established according to a strictly hierarchical structure, i.e., the country is entirely divided into nonoverlapping first-level administrative units, each of these in

turn is divided into nonoverlapping second-level units, and so on. However, there are often other types of units, especially those of an urban character, which do not fit neatly into such a hierarchical structure (see, for example, the discussion of sanitary districts in the Thailand case-study).

In developing a design for an IHSP, it is necessary to have a detailed knowledge of the country's administrative structure: the kinds of units that exist and their relation to one another. This information is relevant to three aspects of design: the data requirements, the sample design and the structure of field operations.

With respect to data requirements, there may be an existing or anticipated need for separate survey estimates for first-level administrative divisions, such as states or provinces, or for individual large cities or metropolitan areas. In addition or alternatively, at the national level, it may be desired to produce separate estimates for the rural population and for the urban population by size of place. Such classifications are usually based entirely or partly on administrative area definitions.

As will be discussed in detail in chapters IV and V, the development of sampling frames and the design of samples are strongly influenced by the administrative structure of the country. This is especially true with regard to stratification and the choice of sampling units at various levels.

In countries that operate with a decentralized field force, i.e., with interviewers working out of local or regional offices, it usually makes good sense to have the area for which each office is responsible coincide with the area of one or more administrative divisions or subdivisions.

A recent population census is usually a good source of information about a country's administrative structure. Census publications provide listings and data for at least the larger units. Unpublished data are often available for lower level units, because of their role in the designation of enumeration areas for the field-work.

The administrative structure of a country is seldom permanently fixed. New units and, occasionally, new kinds of units are created to accommodate changes in the size and distribution of the population and new administrative requirements. It is even possible for existing units to be completely eliminated. In Sri Lanka, for example, entire villages have been inundated and ceased to exist as the result of large-scale hydroelectric projects.

A plan for an IHSP needs to make allowance for such changes. It is therefore necessary to find out which government agency or agencies are responsible for the creation and definition of various kinds of administrative divisions and subdivisions, and to establish a systematic arrangement for those agencies to provide the statistical office with timely information about all changes.

A detailed, systematic account of the different kinds of administrative divisions and subdivisions that exist can be very helpful in the design of an IHSP. If not already available, such an account should be prepared. For an example of such an account, see Skunasingha and Jabine, 1983.

2. Characteristics of the target population

The design and conduct of household surveys are affected in numerous ways by characteristics of the country's target population, such as languages spoken, ethnic backgrounds, religions, cultural practices and principal economic activities. A population that is very diverse with respect to these characteristics may require quite different treatment from one that is more uniform. A full discussion of the implications of these characteristics for survey design is beyond the scope of this study, but there are two specific aspects that relate very directly to frames and sample designs and will be discussed: living arrangements and mobility.

The term living arrangements is used here to refer to the relations of persons, families, households and other basic societal units to the structures -- be they houses, apartments, huts, tents, boats or other types -- in which they live, i.e., their living quarters or housing units. The United Nations (1980a) has established recommended definitions of households and housing units for use in population censuses. In some societies there is virtually always a one-to-one correspondence between households and housing units; however, this is by no means true everywhere. In some countries extended families living in compounds are common. Members of these extended family groups may take their meals in one structure and sleep in another. Within a country, typical living arrangements often vary between urban and rural areas and from one region to another.

The appropriate choices of ultimate sampling units (USUs) and the rules of association to relate the elementary units to the USUs depend to a considerable extent on the kinds of living arrangements that exist. Alternatives are discussed in Chapter IV, Section B. Designers of samples for household surveys need to be familiar with different kinds of living arrangements in their countries in order to make intelligent decisions among these alternatives.

Mobility of the population refers to changes, both temporary and permanent, in living arrangements. At one extreme are very stable populations whose living arrangements seldom change. At the other extreme are nomadic populations that change locations several times each year, or small tribal groups who practise slash and burn cultivation and change their locations every few years. In some areas, many families move from villages to farmland areas for extended periods every year.

The extent of mobility influences various aspects of survey design. Nomadic groups are sometimes excluded from the survey population entirely. If not excluded, they may require special sampling and data collection procedures. If individuals frequently leave their normal places of residence for extended periods for occupational, educational or other reasons, this can affect the household definitions used in surveys, especially the rules that associate persons with households. Greater mobility would favour the use of <u>de facto</u> as opposed to <u>de jure</u> rules. Likewise, greater mobility would, for several reasons, shorten the periods for which master sampling frames and master samples can be used effectively without updating or complete redesign. It would also favour the use of housing units, rather than households, as USUs.

3. Access to sample units

The development of optimum sample designs requires a balancing of sampling efficiency against sample selection and interview costs. For a fixed-size sample of households, sampling efficiency would be best served by dispersing the sample households as widely as possible throughout the target population. However, sample selection and field costs for this kind of design would be much higher than if the sample households were highly clustered. The optimum design lies somewhere between the extremes. Exactly where it lies depends on how much of field-workers' time will be needed to reach and gain access to sample households, as opposed to actually conducting interviews. Time required to "gain access" includes time needed to reach the general location of the sample households, to travel between households and, if necessary, to obtain permission to conduct interviews from local authorities.

Thus, background information needed for design purposes includes: population density in different parts of the country, the extent and quality of the road network, the availability of various types of public transport, and the costs of interviewers using their own or agency transport, such as jeeps, motor scooters or bicycles. Also important are the availability and costs of food and lodging for interviewers who are required to spend one or more nights away from home.

In theory, if a country has had some experience conducting household surveys it should be possible to obtain information on field costs and on the relative amounts of time spent by interviewers on gaining access and on actually conducting interviews. Although these data would be directly applicable only for the designs actually used, they could also be used to estimate unit costs needed for modeling the costs associated with alternate designs. Unfortunately, records of survey interviewing activities usually do not provide sufficiently detailed and accurate information for this purpose; thus, special studies in connection with pretests or ongoing surveys are likely to be needed. A small investment in special record-keeping activities may lead to substantial improvements in design efficiency. Sometimes access to units may be completely lacking, either on a seasonal basis, e.g., because of annual flooding or impassable roads in some areas, or for more extended periods, e.g., due to civil disturbances or security problems. Seasonal access problems can be anticipated and appropriate adjustments made in the scheduling of data collection. Other kinds of access limitations are less predictable and will have to be dealt with as they arise.

4. Availability of sampling materials

Turning now to the inventorying and analysis of resources available for household surveys, the first category to be considered is materials available for use in sampling. Specifically, one needs to know what kinds of lists or maps are available to use for construction of frames at various stages of sampling, and what data are available to use as measures of size for frame units. This topic is covered in detail in Chapter IV: a brief summary is included here to make this section self-contained.

The source most commonly used to construct frames for household surveys is the latest census of population. The census itself requires the development of a frame with units small enough to be covered by individual census enumerators. Census maps of reasonable quality and counts of population or households, if available for these units, can serve as the main basis for the initial development of a master sampling frame. If there has been no recent census or if the needed materials from the census have not been preserved, it will be necessary to look for other sources of maps and lists of potential frame units.

There are many possible sources. Government ministries that administer broad-based programmes relating to education, health, public safety and agriculture are likely to have lists and maps of the administrative units in which programme activities are carried on. Often there is a national agency with primary responsibility for mapping activities and the naming and definition of political divisions and subdivisions. For some areas, especially cities and towns, maps of acceptable quality may be available from local authorities.

The quality of maps and lists obtained from censuses and other sources cannot be taken for granted. Assessments for completeness and accuracy should include field checks and checks against alternate sources of information.

Findings from an inventory and assessment of available sampling materials will be a major factor in deciding what kinds of sampling units to use in constructing a master sampling frame and in designing samples for an IHSP. If the scope and quality of the materials are limited, it may be better at the start to apply resources to the planned development of better sampling materials, rather than proceeding immediately with an ambitious programme of surveys. 5. Field staff

It is unrealistic to plan for a continuing programme of household surveys unless a regular field staff exists or there is a prospect of establishing one. If there is already a regular field staff, it may have responsibilities other than the collection of data for household surveys. To evaluate its potential use in an IHSP ask:

- o How many of the field staff and how much of their time are available for household survey listing and interviewing, and at what times of year?
- o How is the field staff based, i.e., centrally, regionally or locally?
- o To what extent are transport facilities and funds available for field-work away from the home base?
- o What provisions exist for periodic training on the content and procedures for household surveys and for supervision of the field-work?

Answers to questions like these may lead to the conclusion either that additional field personnel are needed or that the scope of the planned surveys must be cut back, e.g., by foregoing subnational estimates or scheduling fewer surveys or survey rounds each year. Managerial judgements will be necessary concerning the prospects of taking on additional staff: how many and when?

If there is no regular field staff available for work on household surveys, a detailed plan, including cost estimates, will be needed for establishing one. Basic decisions will be required as to numbers, basing arrangements, transport facilities, ratio of supervisors to interviewers, interviewer qualifications and pay scales, whether interviewers should be employed full- or part-time and related matters. The plan for the field staff should include provisions for any surveys that may require interviewers with special qualifications.

One might ask whether the structure and size of the field staff should determine the sample designs for planned surveys, or vice versa. Probably the correct answer is "neither of the above": they should be planned together, taking into account the data requirements, the operating environment, and the total resources available for household surveys.

6. Data processing capabilities

Surveys have no value until the results are available to users. Survey results delivered to users three or four years after the data collection are of much less value than those that are made available within a few months of the data collection. Lack of timeliness of results is without question the number one criticism of household survey programmes undertaken by developing countries. Consider the three major stages of a survey: design, data collection and data processing. Various difficulties may be encountered in the first two stages, but they are usually overcome and the work proceeds more or less on schedule. It is in the data processing stage that serious delays occur, negating the accomplishments of the preceding stages.

Survey data processing requires special equipment and skills. These requirements are explained in detail in the NHSCP technical study <u>Survey</u> <u>Data Processing: A Review of Issues and Procedures</u> (United Nations, 1982b). In planning the scope and content of an IHSP, it is essential to make a realistic appraisal of the resources for survey data processing that are or are likely to be available. These include: computer and data entry equipment; stardard software for survey processing; clerical personnel for manual edits, coding and data entry; and systems designers and programmers capable of developing the customized procedures and computer programs that will be needed.

Some basic principles that, if scrupulously observed, will help to avoid overload of data processing capabilities, with consequent delays, are:

- (1) Severely limit the amount of information to be collected on each survey topic.
- (2) When survey topics are repeated, use the same questionnaire formats and processing procedures each time. Exceptions should be made only when it is clear that changes will produce substantial improvements in quality or relevance of the results.
- (3) When multiple topics are included in a single survey or survey round, develop modular processing systems, so that data for topics not requiring direct linkage can be processed independently, according to established priorities.

Each of these guidelines recognizes that survey data processing is a complex activity, much more so than might appear on the surface, and that its complexity is closely linked to the number and nature of data items to be dealt with in a particular sequence of processing operations.

7. Technical and managerial staff

Data processing is not the only aspect of household survey planning and execution that calls for technical expertise. Other specialized skills needed for surveys include:

- o cartography
- o sampling

- o questionnaire development
- o training
- o subject-matter analysis and report preparation
- o publications design

Many of these skills are needed throughout the survey programme. The need for the first four can be especially heavy in the early stages and may require seeking assistance from outside sources. However, as mentioned in section 1 of this chapter, a long-range goal should be to develop in-house staff with the necessary expertise in most of these areas.

Survey management is an especially demanding task. In most national statistical organizations, survey operations require the participation of several different offices or divisions. For good results, it is important that the responsibility for each phase of survey planning and operations be assigned to a lead unit within the organization and that a detailed timetable be prepared for all survey activities. Furthermore, primary responsibility for actual performance should be given to a survey manager who is placed at a high level in the organization and is able to spend full time monitoring performance and to arrange for adjustments when they become necessary.

There may be a temptation to assume that surveys can be carried out successfully provided the necessary field staff and data-processing facilities are in place. However, qualified technical and managerial staff are every bit as necessary and their current or prospective availability must be weighed carefully in deciding on the scope of survey activities that can be undertaken.

D. <u>General classes of designs for integrated</u> household survey programmes

The country designs for IHSPs that are described in the case-studies in Appendix A vary widely. It is evident that there can be no "cookbook" approach to the design of an IHSP. Nevertheless, some general classes of designs can be identified. The purposes of this section are to describe the principal classes of IHSP designs, to give some examples of each, and to discuss the implications of each class of design for the development of master sampling frames and master samples.

IHSPs can be distinguished by the number of separate surveys included in the programme and by whether the individual surveys are multi-subject or single-subject surveys. On this basis, three kinds of IHSP designs can be identified:

- Design B Two or more single-subject surveys. The conduct of individual surveys may be continuous, periodic or ad hoc.
- Design C Two or more separate surveys, at least one of which is a continuing or periodic multi-subject survey.

No category is included for a single survey covering only one subject. While conducting such a survey might be an appropriate first step toward implementation of an IHSP, it would not in itself constitute an <u>integrated</u> programme of surveys, and is therefore outside the scope of this technical study. As stated earlier, the sample design and other aspects of single-subject surveys have been treated extensively in several textbooks and manuals.

Once taken, a decision to adopt one of these three kinds of designs is not inalterable. The Indian National Sample Survey (see Case-Study: India) started with design A, i.e., in each round, data on several topics were collected from the same sample of households. Later, however, this design was abandoned, largely because of respondent fatigue, and design B was adopted. Some countries may start with a single multi-subject survey (design A) and then, as resources and user needs expand, add one or more single-subject surveys, thus switching to design C. The historical development of the system of national household surveys in the United The oldest of the U.S. surveys, the States followed this pattern. Current Population Survey, is a monthly multi-subject survey. A standard labour-force inquiry is repeated each month and supplements on various topics are included on a regular or ad hoc schedule. Later, household surveys on other topics, such as health, housing and crime were added, using separate samples of households but overlapping samples of PSUs and More recently, the samples for these specialized surveys have been SSUs. made fully independent of the sample for the Current Population Survey.

Each of the three principal IHSP designs -- A, B and C -- has a number of variants, which can be identified by examining a few key design features. Common variants for designs A and B are discussed below, with illustrations taken from case-studies and elsewhere. A detailed discussion of design C would be superfluous: its variants are created by combining variants of designs A and B.

1. Design A: a single multi-subject survey

A continuing or periodic multi-subject survey consists of survey rounds, for each of which the data collection period might be as short as a week or two or might last as long as a full year. The survey is designed to produce separate estimates from each round. Basic demographic and socioeconomic variables and survey modules on one or more key topics, such as labour force or agricultural activities, are included in every round. Other topics vary from one round to the next. In a survey with more than one round per year, certain topics might be covered in the same round or rounds every year. Other topics may be included at intervals of more than one year, or on an ad hoc schedule.

Thus, the basic structure of a continuing or periodic multi-subject survey with respect to timing and content is described by:

- o Length of the data collection period for each round
- o Number of rounds per year
- o Topics covered, by round
 - Included in every round
 - Other

Another important dimension of the design of a multi-subject survey involves the distribution of the sample households within rounds and the nature of overlap between the samples used in different rounds.

There are two basic patterns for the distribution of sample households within a survey round. These two patterns are distinguished by whether or not the sample for the round is divided, by some random or systematic process, into subsamples that are assigned to specific time periods within the round (sometimes called sub-rounds). The pattern with no subsampling is most appropriate for periodic surveys in which the rounds are short, say one or two weeks, and data are collected for a fixed reference period. The pattern with subsampling is frequently used in continuing surveys. In a survey round lasting three months, for example, there might be 13 sub-samples, each consisting of households (or areas) for interviewing during a particular one-week period. For some of the survey topics, the reference periods might vary throughout the round. For example, for labour force activity the reference period could be the seven days preceding the interview or the calendar week preceding the interview week.

The estimates that can be developed from surveys using these two different within-round sample structures are not precisely equivalent. To illustrate this, consider a labour force inquiry in a survey with four quarterly rounds. If the complete sample for each round is interviewed within a short period (the periodic approach), estimates of labour-force participation rates for each quarter will refer to a specific week or other short period within the quarter. If the complete sample for a round is divided into weekly subsamples (the continuous approach), estimates of labour force participation rates will be average values for the quarter.

The continuous approach is somewhat more flexible in the sense that data can be accumulated to produce estimates for any desired time period, subject, of course, to limitations imposed by sample sizes. On the other hand, the continuous approach requires dedicated full-time interviewers, whereas the periodic approach permits use of interviewers for other activities in the interim periods.

Sample overlap <u>between</u> rounds is characterized by the stages of sampling at which it occurs and by the proportion of sample overlap at each stage. If samples for successive rounds are selected entirely independently of each other, some overlap will usually occur as a matter of chance. If the selections are not independent, the proportion of overlap can be controlled at any desired level from 0 to 100 percent.

Some simple illustrations may clarify these ideas:

<u>Example 1</u> - 100 percent overlap at the first stage, chance overlap at subsequent stages. The same sample of PSUs is used for every round. Sampling within PSUs (one or more stages) proceeds independently for each round.

Example 2 - No overlap at any stage. A large master sample of census enumeration areas is selected and divided into subgroups, using a systematic procedure. A different subgroup is used for each survey round. Regardless of the method of sampling within PSUs, there is no overlap from round to round.

Example 3 - 100 percent overlap at the first stage, 75 percent overlap between rounds at subsequent stages. The same sample of PSUs, e.g. administrative districts, is used for every round. Census enumeration areas are used for the second stage of sampling; for each round, one in four of the sample enumeration areas from the previous round are replaced by new enumeration areas. A sample of housing units is selected in each sample enumeration area and is retained as long as that enumeration area remains in the sample.

Other, more complex patterns of overlap are possible. In a survey with quarterly rounds, for example, it would be possible to design a sample with 50 percent overlap between adjacent quarters and 50 percent overlap between quarters one year apart. This could be done by introducing a new subsample of households each quarter, to be retained in the following quarter, left out for the next two quarters, and then brought back into the sample for two more quarters (see Exhibit 3.4).

Year and round								
Subsample	Rl	Year 1 R2	R3	R4	R1	Year 2 R2	2 R3	R4
1	x							
2	X	Х						
3		Х	X					
4			X	Х				
5	Х			Х	X			
6	Х	X			х	X		
7		Х	X			Х	Х	
8			X	X			х	X
9				X	X			Х
10					X	х		
11						Х	х	
12							Х	Х
13								X

Exhibit 3.4 A design with 50 percent quarter-toquarter and 50 percent year-toyear overlap

Strictly from the point of view of sampling efficiency, sample overlap from round to round offers two advantages: better estimates of change and reduction of the costs associated with the introduction of new sample units. On the other hand, if it is desired to accumulate estimates over more than one round, the most efficient design is the one that introduces an entirely new sample for each round.

The application of the master sample concept to a continuing or periodic multi-subject survey is straightforward. It consists in the selection, at some stage, of a sample from which different subsamples will be selected for use in different survey rounds. Two illustrations from the case-studies follow:

Example 1 - Jordan

A periodic multi-subject survey was planned, with two rounds per year in the first two years and four rounds per year in years three to five. Before the start of the survey, 21 independent replicate samples were to be selected, each consisting of 50 area segments (blocks or block groups) with designated subsampling rates. One or more of these replicates would be used for each round of the survey, with an unspecified pattern of overlap between rounds.

Example 2 - Saudi Arabia

A plan was developed for a multi-subject survey, to be conducted in quarterly rounds over a five-year period. A two-stage self-weighting sample of segments, each containing from 100 to 200 households (expected size) was to be selected. A listing of households was to be prepared for each segment and each household was to be randomly assigned to one of eight subsamples. In each of the four quarterly rounds for the first year, subsamples 1 through 4 were to be used. For the second year, subsample 1 was to be replaced by subsample 5, and so on, through the remaining three years.

In the Jordan example, the primary benefit of using the master sample technique is the efficiency gained by selecting, in a single operation, all of the area segments needed for a five-year period. In the Saudi Arabia example, it was apparently intended that the household listings prepared initially for each of the sample segments would provide the samples of households needed over the five-year period of the survey. In practice, one would expect that a procedure would be needed to update the listings, in order to avoid biases resulting from changes in households.

To summarize this discussion of multi-subject survey designs, the basic design features that are most relevant to the use of master sampling principles are shown in Exhibit 3.5.

2. Design B: two or more single-subject surveys

IHSP designs consisting of two or more single-subject surveys can be characterized by the kinds and extent of integration among the surveys that make up the programme. It is convenient to define three levels of integration:

- Level 1. Sharing of facilities, including field staff, processing staff, processing equipment and master sampling frame.
- Level 2. Use of the same master sample for more than one survey.
- Level 3. Deliberate inclusion of the same housing units or households in more than one survey.

All of the IHSP designs described in the case-studies in Appendix A exhibit integration with respect to the facilities included in level 1. Use of the same master sampling frame is a universal feature; this points to the desirability of devoting a careful planning effort and adequate resources to the design, development and maintenance of a master sampling frame (see Chapter IV).

Use of the same field staff for different surveys is common, but not necessarily universal. The upper-level management of the field staff will, of course, be responsible for the field work in all of the

Exhibit 3.5 Basic design features and options for multi-subject surveys

Υ.

CONTENT	Constant (all rounds)					
	 Basic demographic and social variables Core topics 					
	Variable by round					
	 Topics included at regular intervals Ad hoc topics 					
FREQUENCY OF ROUNDS	Number of rounds per year					
TIMING OF INTERVIEWS	Option 1, periodic. Short interview period, no subsampling.					
	Option 2, continuing. Interviewing continuous over round, subsamples designated for sub-rounds.					
OVERLAP BETWEEN ROUNDS	Stages of sampling at which it occurs					
	Proportion of overlap at each stage					
	 Controlled, 0 to 100 percent Uncontrolled, expected proportion depends on sample design 					

.

conducted by the same organization and this household surveys multi-survey responsibility usually extends to the first-level supervisors of the survey interviewers. The documentation on which the case-studies are based is not always explicit about the assignment of interviewers to different surveys, but gives the impression that in most programmes the same interviewers are used for all of the household This arrangement requires a scheduling of interviews for the surveys. different surveys that avoids conflicts and spreads out the work-load evenly.

The case-study for Nigeria provides an illustration of inefficiencies resulting from an uneven distribution, over the year, of the interviewing workload in rural PSUs. For most of the household surveys in each PSU, all field-work, including household listing, was scheduled to be completed within a two-month period. However, price data collection and interviewing for an agricultural survey were spread out over the year and for this reason it was considered necessary to station a permanent team of two interviewers in each rural PSU.

Integration of separate surveys at levels 2 and 3 goes beyond the sharing of facilities. At level 2 the same master sample is used in two or more surveys, but with no overlap (except that which might occur by chance) in the USUs for the different surveys. At level 3, the same USUs are deliberately used in more than one survey.

To illustrate some of the options that exist at levels 2 and 3, let us consider an IHSP with two or more single-subject surveys, all using a two-stage sample design characterized by:

- Use of census enumeration areas as PSUs.
- Listing of housing units or households in sample PSUs.
- Subsampling of housing units or households in sample PSUs.

The main options for sample-design integration in this situation are:

- Option (1). Use of a different sample of PSUs for each survey, selected from the same large master sample of PSUs.
- Option (2). Use of the same PSUs from the master sample in different surveys, but selection of different housing units or households within these PSUs for each survey.
- Option (3). Use of the same sample of housing units or households in each survey. A sub-option under this option would be to use a subsample of housing units or households from one survey as the sample for a second survey requiring a smaller sample than the first one.

All three of these options use the same master sample for different surveys (level 2 integration). Option (3) uses the same housing units or households for different surveys (level 3 integration).

What are the advantages and disadvantages of these three options? Under option (1), the overall cost of sample selection for the set of surveys can be reduced somewhat by selecting the PSUs needed for all surveys. i.e., the master sample of PSUs, in a single operation. However, the number of PSUs needed in the master sample would be larger than that needed under options (2) and (3), and any costs that vary according to the number of PSUs, such as preparation of maps and listings, household would be greater in the same proportion. Furthermore, under options (2) and (3) interviewers could be stationed in the fixed set of PSUs and work only in those PSUs, whereas this would not be possible under option (1).

One would not expect much difference between options (2) and (3) with respect to cost. Use of option (3) would open up the possibility of substantive integration of data from different surveys at the micro (household) level. On the other hand, using the same sample of housing units or households for different surveys would increase the burden on these units, which could have a negative effect on the quality of results.

The three options just described do not exhaust the ways of integrating samples for different surveys. Another possibility is to use household listing in sample PSUs as a screening device to locate households with specific characteristics. One example of this has been in the annual migration survey conducted in the Bangkok Metropolis of Thailand. During the annual listing of households for the labour force survey in sample blocks and villages, questions are asked about in-migration of household members during a specified time period. All households with one or more in-migrants are interviewed in the migration survey.

In countries that use more than two stages of sampling, the master sample might consist of second-stage or even third-stage area units. If there is listing and subsampling of housing units or households within these sample units the same three options described for the two-stage design are available.

The case-studies in Appendix A illustrate several different ways of integrating the sample designs for separate surveys. Three examples follow:

Example 1 - India

A new master sample of PSUs (census blocks and villages) is selected annually. Household listings are prepared for the sample PSUs and are used as the frame for separate samples selected for surveys on different topics.

Example 2 - Morocco

The master sample, which was intended to be used for a 10-year period, was to be an area sample selected in two stages. For each of the master sample SSUs, a map was to be prepared showing the division of the SSU into well-defined area segments (tertiary sampling units) with about 40 to 50 households. The general procedure for using the master sample to select samples for different surveys was: (1) From all or a subset of the master sample of SSUs, select a sample of segments, (2) Prepare household listings for the selected segments, and (3) select samples of households from the listings.

Example 3 - Ethiopia

A master sample of 500 rural administrative units called farmers' associations was selected for use over a three-year period. Household listings were prepared for all 500 PSUs at the start of the period and new listings were prepared two years later. These listings were used as sampling frames for selection of households in seven separate surveys. There was considerable overlap in the samples used for different surveys. For example, there was essentially complete overlap between the samples for: the current agricultural survey; the income, expenditure and consumption survey; and the labour force survey. The same sample PSUs were also used for data collection activities other than household surveys, including price data collection and a survey of community-level variables.

3. Choosing an appropriate design

The examples in this chapter and in the case-studies in Appendix A make it clear that possible variations in IHSP designs are virtually unlimited. How, then, can a national statistical office develop a design suited to its particular data needs, operating environment and resources?

To return to the theme with which this chapter began, the primary requirement is to develop a detailed multi-year plan for an IHSP. The nature of the plan will depend on what household surveys are currently being undertaken. If there is none at all, it may be good strategy to start the programme with design A, a continuing or periodic multi-subject survey. Other surveys can be added to the programme later, as needs and resources dictate. If some household surveys are already being conducted, the goal of the plan should be to achieve better integration of these surveys, both with respect to each other and with respect to other statistical programmes, especially population censuses.

The development of the plan should be guided by three objectives: efficiency, quality, and timeliness of results. For the first two objectives, the nature of the field staff that exists or will be established is a basic consideration. For the third objective, timeliness, data processing capability is likely to be the critical factor.

The planning process should be guided by realism about the number and size of surveys that can be conducted with the resources available. An overambitious programme of surveys at the start can overload the available staff and facilities and hinder the longer-range successful development of the programme.

Planning cannot cease once the initial plan is developed; it must be a continuing process. Careful documentation of the cost and quality of survey operations will provide the basis for periodic assessment and modification of the IHSP design.

For countries that are beginning an IHSP, there are two very important steps to be taken once a broad design has been agreed on. The first of these two steps is the development of a master sampling frame that can be used for all of the household surveys. The second step is to decide what kind of master sample, if any, will be used in the programme and, if there is to be a master sample, to design and select it. These two activities, which are closely related to each other and to the broad IHSP design, are the subjects of chapters IV and V, respectively.

CHAPTER IV

SAMPLING FRAMES

Sampling frames are lists of units from which survey samples are selected. For surveys with multi-stage sample designs, a frame is needed for each stage of selection. Frames are also needed for censuses. In a census, all of the frame units (or at least all that meet the specifications for census coverage) are included in the data collection phase.

The choice of suitable frames for all stages of sample selection is a critical aspect of design for IHSPs and for individual household surveys. The cost of developing entirely new frames is likely to be considerable; it is therefore desirable to use survey designs for which the sampling units are covered by existing frame materials, at least for the first stage of sampling. Choices among sample design options should generally favor those options for which acceptable frame materials are more readily available.

The first three sections of this chapter are general in the sense that they cover sampling frames used at all stages of sampling in household surveys. Section A discusses the basic considerations in the choice of sampling frames: intended uses, kinds of units, coverage, media, and contents, including auxiliary materials. Section B focuses more closely on frame units and on the rules of association that link them with each other and with elementary units in survey target populations. Section C discusses desirable properties of frames and includes a check-list for use in assessing the suitability of existing frames and frame materials.

Section D describes procedures for the development of master sampling frames, i.e., frames that list the primary sampling units (PSUs) from which first-stage samples are to be selected. A master sampling frame will normally be the starting point for sample selection in all surveys that are part of an IHSP; therefore its development and maintenance deserve special attention.

For some survey designs, a master sampling frame can be used for one or more stages of sampling beyond the first stage. At some stage of sampling, however, it will almost certainly be necessary to develop special frames for the next stage of selection, e.g., lists of housing units, households or small area segments. These secondary sampling frames, which are needed only for the sample of units selected at the preceding stage, are the subject of section E, the final section of this chapter.

The development and use of frames for individual surveys has been dealt with extensively in the literature on sample surveys. The World Fertility Survey <u>Manual on Sample Design</u> (WFS, 1975) provides an excellent treatment oriented toward household surveys of fertility in developing countries. The treatment of frames in this chapter focuses on the development and use of frames for household survey programmes, as opposed to individual household surveys. In this context, the key topic is the development of master sampling frames that can be used in more than one survey.

A. Basic considerations in the choice of sampling frames

Key considerations in the choice of sampling frames, regardless of the stage of sampling for which they are used, include the following:

KEY CONSIDERATIONS IN THE CHOICE OF SAMPLING FRAMES

- o Intended uses o Units
- o Coverage
- o Media
- o Content
- o Auxiliary materials

Each of these considerations is discussed in this section.

1. Intended uses

Sampling frames are used for <u>sample selection</u> and for <u>making</u> estimates based on sample data.

A sampling frame is required for each stage of selection in a multi-stage design. Sometimes the same frame can be used for more than one stage of selection. For example, a frame consisting of a listing of census enumeration areas (EAs), ordered by province and by district within each province, could be used to select a sample of districts and then, within each of the sample districts, a sample of EAs.

A frame of some sort is also required for a population census; however, in that context it would not be called a <u>sampling</u> frame. Similarly, in some surveys, the USUs are small area segments for which all households living in the segments are to be interviewed (take-all segments). Lists of households or housing units for these take-all segments would not be considered <u>sampling</u> frames. However, in designs for which the USUs are housing units or households sampled from listings for census EAs or other area segments, these listings are the sampling frame for that stage of sampling.

The choice of the sampling method to be used at each stage of selection is limited by the information available for each frame unit at that stage. If the information consists only of attributes (e.g., urban/rural classification, identification of higher-level units), it is necessary to use an equal probability selection method with or without stratification. However, if quantative information (e.g., counts of persons or households from a recent census) is available for all or virtually all frame units, this information can be used in connection with sample selection or estimation, or both. A sample of EAs could be selected with probability proportionate to size (e.g., census population counts). Alternatively, a sample of EAs could be selected with equal probability and the quantative data used for ratio or regression estimation.

A critical aspect of intended use is whether the sampling frame is to be used only in a single survey or in more than one survey or survey round. Frames that are to be used more than once must meet more stringent requirements, especially if the intended uses will occur over a relatively long time period.

2. Frame units

Strictly speaking, the frame units are the sampling units included in the frame. Other kinds of units are often used as "building blocks" from which to construct the frame units.

To illustrate this distinction, consider the process of constructing a sampling frame for an urban area for which the following materials are available:

- o A map which divides the area into blocks with defined boundaries.
- o A current count of housing units for each block.

For sampling purposes it is desired to divide the entire area into segments with housing unit counts of 50 or more. To accomplish this, blocks with fewer than 50 housing units are combined with adjacent blocks. In this example, the segments consisting of one or more blocks are the true frame units, i.e., the sampling units from which a selection is to be made. The blocks are the units from which the segments are derived. The blocks themselves are used as segments whenever this can be done in conformity with the specified minimum size of segment.

A frame can include more than one kind of sampling unit. In the same example, suppose the urban area in question is divided into a large number of administrative units such as wards or precincts and that blocks do not cross ward or precinct boundaries. Segment and housing unit counts could be accumulated for these administrative units and the frame could then be used in three different ways (depending on the survey objectives and the nature of the subsequent stages of sampling to be carried out):

- (1) To select a one-stage sample of wards or precincts.
- (2) To select a one-stage sample of segments, with stratification by wards or precincts, if desired.
- (3) To select a two-stage sample with wards or precincts as PSUs and segments as SSUs.

For the two-stage sample, the creation of segments would be necessary only for the wards or precincts falling in the sample.

The kinds of units in frames used for household surveys include:

- o Area units
 - Administrative subdivisions
 - Census enumeration areas
 - Other
- o Non-area units
 - Housing units
 - Households
 - Persons
 - Other, e.g., nomadic tribes, institutions, construction camps

Area units cover specified land areas with defined boundaries. The boundaries may be physical features such as streets, roads, railroads and rivers, or they may be imaginary lines (shown on maps) representing the official boundaries between administrative subdivisions.

Area units may be administrative units of the country established for governmental purposes, or they may be units established solely for use in censuses or surveys.

Some administrative subdivisions do not, strictly speaking, qualify as area units, for example, the villages used as rural frame units in Thailand. In general, the village boundaries are not officially defined, on maps or otherwise, but it is considered to be relatively easy for field-workers to identify the housing units or households associated with most villages by consulting local authorities. Hence, they are deemed acceptable for use as frame units.

Census enumeration areas (EAs) are usually established within the smallest type of administrative unit that exists in a country, so that EA counts can be added to obtain counts for the administrative units. Another objective in establishing EAs is to limit and more or less equalize the workloads of individual census enumerators.

In some countries, area units smaller than census EAs are established solely for use in household surveys. These units are used mostly for sampling within PSUs or SSUs and are usually established only for sample PSUs or SSUs. One example would be the sample areas or area segments established within the sample "count units" of the United States master sample of agriculture (see case-study).

The commonest types of non-area frame units are housing units and households. Frames consisting of lists of housing units or households

are often prepared for the final stage of sampling within sample PSUs or SSUs. The relative merits of housing units and households as frame units are discussed later in this chapter.

In some household surveys, persons are the USUs and are sampled from listings prepared for sample households. An example occurred in the United States National Health Interview Survey in 1983. For a set of supplemental questions on use of alcohol and tobacco, one or more persons were selected in each sample household, using a random selection process. Lists of persons living in selected institutions are often used as sampling frames in household surveys that include a sample of the institutional population.

There have also been surveys in which listings of eligible persons, e.g., women of child-bearing age, have been prepared for sample segments, without regard to household structure. The World Fertility Survey for Senegal is one example.

Lists of housing units, households and persons are used as frames for the last or next-to-last stage of sampling in multi-stage household samples. Other types of non-area units are occasionally used as PSUs, for example, villages in the rural part of Thailand and farmers' associations in rural Ethiopia (see case-studies).

3. Coverage

This subsection is concerned with the coverage <u>objectives</u> of frames; the actual completeness and quality of coverage are discussed in Section C of this chapter.

The coverage objective of the frame or frames used for a survey is to provide access to all of the elementary units in the survey population and to do so in such a way that every one of those units has a known (or knowable) probability of selection in the sample for the survey. Access is achieved by sampling from the frames, usually through two or more stages of selection and by the use of rules of association that link elementary units to the units that were selected at the final stage of selection, i.e., the USUs.

The frame or frames for the first stage of sampling must provide 100 percent coverage of the designated PSUs. At subsequent stages of selection, frames are needed only for the sample units selected at the preceding stage.

Many countries use separate first-stage frames for urban and rural areas. While these frames could be treated as components of a single national frame, it is usually more convenient to describe them separately as the urban frame and the rural frame. For some surveys special supplementary frames may be needed, e.g., lists of nomadic tribes, lists of institutions or special dwelling places (see case-study for Australia) and, for sample institutions or special dwelling places, lists of inmates or residents. For household surveys of agricultural activities, it may be necessary to develop and use list frames to cover large holdings that are not associated in any direct way with specific households or housing units. An example is described in the case-study for Ethiopia. Sometimes special frames of these kinds are developed for the entire country; in other cases they are compiled only for sample PSUs or USUs.

Some household survey programmes do not attempt to cover the entire civilian non-institutional population of the country. Nomadic groups are specifically excluded from coverage in Ethiopia and Jordan. To date, the household surveys in Ethiopia have covered only rural households. The national household surveys in Thailand have excluded groups known as hill tribes who are not governed under the normal administrative structure; however, there have been special surveys of the hill tribe people. Naturally, these kinds of exclusions are reflected in the nature of the frames constructed for those countries.

4. Media

Sampling frames may be stored either on print (hardcopy) or on electronic media. Electronic media, such as computer tapes or disks, if used at all, are most likely to be used for the master sampling frames that are needed for the initial stages of sample selection.

For a frame stored on an electronic medium, it is relatively easy to produce a printout of the entire frame or of any portion desired. Conversely, a frame existing only in hard-copy form can be transferred to an electronic medium. The cost may be substantial, but if the frame is expected to be used for complex sample selection procedures or for several different samples, the additional expenditures may be justified. Computerization of a master sampling frame can provide flexibility in the choice of sample designs and can eliminate lengthy clerical operations which might otherwise be needed to group frame units into new strata, to select sample units with probability proportionate to size, or to generate combined measures of size based on two or more variables.

5. Content

The frame contains a record for each frame unit. What information should this record contain?

The only item that is absolutely indispensable is a <u>unique</u> identifier of each unit. If a unit is selected, the identifier provides the means of access to the unit in order to perform subsequent sampling operations or to collect survey data.

Names of administrative units, such as villages, are not necessarily unique within a country. To ensure uniqueness and to facilitate sample selection and control operations, a carefully designed system of numerical identifiers is needed. Desirable properties of such a system are discussed in subsection C, 1. The numerical identifiers will, of course, be linked with other identifiers, such as place names or addresses of housing units, either in the frame itself or on maps or other auxiliary materials.

Naturally, a frame can be used much more effectively and efficiently if its unit records contain some information in addition to the primary identifiers. Exhibit 4.1 shows eight categories of information that can usefully be included on records for frame units. The first four categories shown in the exhibit all fall under the general heading of identifiers: they provide a unique identification of each unit, help to locate it, and link it to higher level units. Categories 5 and 6 cover information used in sample designs that are more efficient than simple random sampling. The same information may also be used in connection with sample estimation. The last two categories, 7 and 8, reflect the use and updating of the frame. These kinds of information, of course, are only necessary if the frame is to be used more than once.

Frames do not need to include all of these categories of information. However, it is usually easier to include items for which the need is not obvious when the frame is initially developed than it is to incorporate them later. Also, it would be a sensible precaution at the start to leave a few blank areas in the paper or computer records to add information for which the need is not initially foreseen.

6. Auxiliary materials

Maps seldom serve directly as sampling frames. Even when the frame units are areas defined on maps, it is customary to make a separate listing of these units to use as the sampling frame proper.

Maps, however, are an essential adjunct to most sampling frames. They serve several important purposes. Small-scale master maps show the general location of area frame units, such as census enumeration areas, blocks or villages, within larger administrative subdivisions. Such maps help field staff to locate the area units assigned to them. They can sometimes be used for other purposes, such as assigning numeric identifiers to frame units in a sequence based on physical proximity. Large-scale maps or sketches show the boundaries of frame units that are to be listed or subdivided into smaller area units. They can also be used to record the location of individual housing units or dwelling units, so that interviewers can more readily locate specific units assigned to them for interviewing.

Other useful auxiliary materials include: summary tabulations of frame content, e.g., distribution of frame units by administrative subdivision, by size and by selected stratification variables; records of units selected for specific samples (if not recorded in the frame itself); and, for computerized frames, record layouts with definitions and source information for each record item.

at may be included	Examples, remarks	Necessary when frame includes mor than one type of unit, e.g., districts and villages within districts	Preferably numeric	Names, addresses, keys to maps	Necessary for subnational estimates. Some of this information may be incorporated in primary identifiers	Urban/rural, city size, major economic activity, population density, etc.	Census counts of population or households, etc.	Needed to avoid unnecessary burden on particular units	E.g., in a housing unit listing, to show which units have been added in an update
.l Categories of information the in records for frame units	Purpose(s)	Distinguish different types of frame units	Uniquely identifies frame unit	Aid in locating unit	Identify higher-level frame units and administrative subdivisions of which unit is part	For grouping of units prior to sample selection	For use in PPS selection, stratification, estimation	Show which units have been selected and for which surveys	Keep track of nature and sources of changes
Exhibit 4.	Information categories	Type of unit code	Primary identifier	Secondary identifiers	Links to higher-level units	Stratification variables	Measures of size	Sample usage indicators	Change indicators
		.:	2.	Э	4.	5.	.9	7.	8.

B. Frame units and rules of association

When a multi-stage sample design is used, a frame is required for each stage of sampling. The frame at each stage is a list of sampling units from which the selection is to be made; thus the frame units are, in fact, sampling units.

Rules of association are needed for two purposes:

- o After each stage of sampling except for the ultimate or final stage, to determine which of the sampling units to be used for the next stage are associated with the sampling units that have been selected in the current stage.
- o After the ultimate stage of sampling, to determine which of the elementary units in the target population are associated with the USUs that have been selected.

Consider, for example, a straightforward three-stage sample design:

Illustrative Design

Stage	Sampling Units
1	District (PSU)
2	Village (SSU)
3	Housing unit (USU)

After the stage 1 sample of districts has been selected, it will be necessary to associate or link specific villages (SSUs) with the selected districts (PSUs). This will be a simple matter if there is a complete list of villages and every village is located entirely within a single district; otherwise a more complex rule of association may be needed.

Following the stage 2 selection of villages, a rule will be needed to link housing units (USUs) with the sample villages (SSUs). This rule will be used to guide field staff responsible for listing housing units in the sample villages. If good quality maps showing the boundaries of sample villages are available, the rule will be straightforward; if not, a more complex rule may be required.

The stage 3 selection will result in a sample of housing units for each sample village. The rule of association used following this stage of selection will guide interviewers in identifying the elementary units -- households and persons -- that are associated with the sample housing units and for which data are to be obtained in interviews.

The purpose of this section is to provide a systematic and comprehensive treatment of the usual practices and the problems that may occur in developing and applying these two kinds of rules of association. The choice of a suitable rule is sometimes obvious, but often it is not. Therefore, the specification of association rules to link sampling units between stages and elementary units with USUs should be made an explicit part of the survey design process.

The basic requirement of a rule of association is that it must give every sampling unit or elementary unit a probability of selection that is not zero and can be accurately determined. Returning to the previous illustration, suppose that each village in the country lies within a single district and that a complete listing of villages has been developed, at least for the sample districts. The rule of association of villages with districts is obvious, and the overall probability of selection of any village will be the product:

Probability of	Probability of
selection of X	selection of the
the district	village within
	the district

If it is possible for a village to be split between two or more districts, the choice of a suitable association rule is somewhat more difficult. The usual and preferred practice is to use a rule that associates each village with one and only one of the districts in which it is partly located. This type of rule can be called a <u>unique</u> rule of association. Specifically, in this case, the rule might be to associate each split village with the district containing the largest part of its population. Alternatively, area might be used as a criterion, depending on which kind of information is more readily available. In case of ties, i.e., half of the population or area is in each of two districts, the rule might be to associate the village with the district having the lower serial number.

If villages and districts have direct administrative links, the best rule might be to associate each split village with the district to which the village chief reports administratively.

The result of such unique rules will be that each split village will be given a chance of selection at stage 2 if and only if the district with which it is uniquely associated has been included in the sample at stage 1.

It is theoretically possible to devise unbiased <u>multiple rules of</u> <u>association</u>, i.e., rules that are not unique in the sense just described. Using the same illustration, each split village could be given a chance of selection if any of the districts in which it is located were selected at stage 1. The within district (conditional) selection probabilities for each split village could be adjusted to allow for the fact that it could be selected from more than one district. However, multiple association rules of this kind are not normally used in household surveys and their use cannot be recommended. They lack a property which is desirable, if not essential, in choosing rules of association: simplicity or ease of application.

Going on to the next stage of our illustrative design, what can be said about the association of housing units with villages? If good quality maps showing the village boundaries are available, the rule of association (which is in the form of instructions for listing housing units in sample villages) is simple: list all housing units located inside the village boundaries. In theory, it is possible for a housing unit to be partly located in two villages, but this is likely to be such a rare event that it would probably be better not to complicate the listing instructions by including a special provision to deal with this possibility.

What if good village maps are not available? In this situation it is more difficult to establish a clearcut rule of association. The objective is obvious: to list every housing unit "in" the village. The problem is to define operating procedures that will ensure that listers meet or come close to meeting the objective. In some countries or regions village officials will have complete or fairly complete listings of houses or families. A requirement that the lister prepare a sketch map showing the locations of individual housing units may lead to more systematic coverage of the villages by the lister and the sketch maps will also help interviewers to locate sample households. Listers can be instructed to inquire, when working near the presumed boundaries of the village, about other houses in the vicinity and to determine whether they are part of the same village. In addition to operations carried out by the listers themselves, the procedures should include some supervisory checks on the completeness of listing. What all of this amounts to, then, is an implicit rule of association and some procedures for implementing it.

Going on to the final stage of our illustrative design, a rule of association is needed to link households (and hence the members of those households) with sample housing units. Part of the rule consists of a precise definition of the term "household". This definition tells which persons should be included in each household: how to treat persons who are usually present but are temporarily absent and those who are temporarily present but usually live somewhere else. A detailed discussion of the alternative <u>de jure</u> and <u>de facto</u> definitions can be found in the <u>Handbook of Household Surveys</u> (United Nations, 1984). The definitions will also explain the circumstances under which persons living in close proximity will be treated either as a single household or as two or more separate households.

In our illustrative design, the rule of association must uniquely determine which of the existing households are associated with the sample housing units. If a sample housing unit contains one or more complete households, a straightforward rule is to include all of them in the sample. It is also possible for a household to occupy more than one housing unit, e.g., if household members occupy two or more housing units in a compound but take all of their meals together in the same place. Probably the best association rule for this situation is to link the household with the housing unit in which the head of the household normally sleeps. With this rule, the household would be included in the sample only if that particular housing unit had been selected.

The foregoing discussion of the rules of association needed for a simple three-stage design has served to illustrate the general nature of such rules and to provide solutions to the questions that occur most often. Two other questions will now be considered: the linkage of household enterprises with households and the kinds of association rules that may be needed to deal with changes over time in sampling units or elementary units.

Household surveys are sometimes used to collect data on agricultural holdings or nonagricultural household enterprises. For this purpose a rule of association is needed to determine which enterprises are linked with households in the sample. Some enterprises are operated in partnership by two or persons. nore This does not cause anv complications if all persons involved are members of a single household, but if they live in different households an association rule is required to decide when to include the enterprise in the sample. What is needed is a rule that will unambiguously associate each enterprise with a single person, so that the enterprise will be included in the sample if and only if the household of which that person is a member is included in the sample. Various such rules are possible, e.g., associate the enterprise with (a) the partner considered to be the senior partner, or (b) the oldest partner, or (c) the partner whose surname comes first in the alphabet. Each of these possibilities leads to an unbiased rule of association; the first might be preferred on the grounds that the senior partner should be best able to give an accurate account of the enterprise's activities.

Changes over time in sampling units or elementary units can be somewhat difficult to cope with, especially in the context of a master sample in which the sampling units at some level are to be used in several surveys over an extended period. Changes in administrative subdivisions, especially at the lower levels, occur in most countries. What should be done when the administrative subdivisions affected have been used as sampling units? If many changes have occurred, it may be necessary to construct a new frame and resample, either for the entire country or for the areas in which there have been extensive changes. This can be time-consuming and costly, however. A feasible alternative, if there have not been too many changes, is to develop unbiased rules of association to link old and new units. Returning to the first stage of our illustrative three-stage design, suppose it is found that there have been changes affecting the number and structure of the districts. These changes might be of three kinds:

- (1) A district has been split to form two or more new districts. One unbiased rule for this case would be to include all of the new districts in the sample if the old district had been included.
- (2) Two or more districts have been combined to form a single district. Two possibilities can be considered. One would be, in effect, to ignore the change and associate the land area contributed from each of the old districts with that district. However, if it were desired for some reason to retain full districts as PSUs, the new district could be associated with the old district that contributed the most population (or area) to it. The other old district would be treated as though it had no population.
- (3) More complex changes, e.g., parts of two districts have been combined to form a third district. As in case (2), at least two options are available. One is to ignore the change, i.e, to retain the two original districts as sample units. The other is to associate the new district with the old district that contributed the most population or area. The second option, with association based on area, is illustrated below.

Districts A and B prior to change (boundaries of new district shown by broken lines)

Areas associated with old districts A and B subsequent to change. The new district and the remainder of old district B are combined to form B'.





Association rules of this kind do not affect the selection probabilities for samples that have already been selected: unbiased estimates can still be made using data for the redefined sampling units with their initial selection probabilities. It should be kept in mind, however, that changes in administrative subdivisions are often made in response to substantial changes in the population of the areas affected. If this is the case, the use of such association rules, while unbiased, may cause significant increases in sampling variance (the betweendistrict component of the variance in the illustration). For this reason it may be necessary to consider other ways of dealing with the changes, e.g., creating a separate "growth stratum" in the master sampling frame and selecting new master sample PSUs from that stratum. This problem will be discussed further in connection with the updating of master sampling frames and master samples.

Changes also affect list sample units, such as housing units and households. Housing units may be demolished or converted to nonresidential use. A single housing unit may be converted into two or more units. New units are constructed.

Households, which are persons or groups of persons living together, are, in most societies, even more prone to changes than are housing units. Household changes occur for reasons such as births, deaths, marriages and work-related moves. Most housing units are fixed structures, but any household can move to a new location close to or far away from its previous site.

How do such changes affect the use of list sample units in surveys? Changes can occur in the interval between the time a listing of housing units or households is prepared and the time it is used for sample selection. When sample units are used for more than one survey or survey round, changes may take place between the initial use and subsequent uses.

Rules of association can be used to some extent to deal with changes in list units in an unbiased way. In particular, simple rules can be established to deal with changes in <u>existing</u> housing units. If a housing unit is demolished, it can be treated as a zero observation. If a unit is split to form two units, both units can be associated with the original unit. If two units are combined, the new unit can be associated with the one having the lower serial number. Rules of this kind do not, however, account for newly constructed units or units converted from nonresidential use. Some kind of procedure for updating listings is needed to identify these new units.

It would also be possible, in theory, to develop unbiased rules of association that link households existing at two different times. However, such rules have two drawbacks. First, they must be complex in order to deal with all possible changes in the structure of households. In practice, the linkage rules would have to be applied in the field and the accuracy of their application would be hard to control.

A second and more serious problem is that household changes usually involve moves over varying distances. A sample design requiring that listed or sampled households be retained regardless of location would substantially increase the cost and complexity of field operations.

One must conclude that listings and samples of housing units and households are perishable and should not be used over extended periods without updating or adjustment. The useful life of a listing or sample of housing units can be extended by the use of suitable rules of association in conjunction with a procedure for identifying new units, but the same cannot be said for a listing or sample of households. Therefore, households should not be used as USUs in a master sample designed for use in surveys or survey rounds taking place at different times.

To summarize the points covered in this section:

- o Rules of association serve two purposes: to link lower-level sampling units with higher-level sampling units and to link units of observation with USUs.
- Any rule of association chosen must ensure that the selection probability for each of the lower-level units to be linked can be calculated precisely and is not zero.
- o Simple rules are preferred to complex rules. Unique rules of association (linking each lower-level unit to one and only one higher-level unit) are preferred to multiple rules.
- o The effects of proposed rules of association on sampling errors need to be considered.
- o In the context of master sampling frames and master samples, rules of association can be used to link initially-existing units with those existing at subsequent times.
- Housing units are more stable than households and should therefore be preferred to households as USUs in listings or samples intended for use in more than one survey or survey round at different times.

C. Desirable Properties of Frames

Frames for household surveys can be constructed in many different ways, using materials from various sources. The purpose of this section is to help survey designers to make choices among alternatives. As in the previous sections of this chapter, the discussion is general: it covers both master sampling frames and frames that are constructed only for sampling units that have been selected for inclusion in one or more surveys. The relative importance of different frame properties will, of course, vary according to the particular purpose for which the frame is intended, and this will be taken into account in the discussion.

To guide the discussion, Figure 4.2 provides a checklist of desirable frame properties. The properties shown in the checklist have been
grouped in three major categories: properties related to quality, those related to efficiency, and those related to cost. In using the checklist as a decision aid, the interrelationship of these categories must be considered. Development of a frame that ranks high on both quality and efficiency of use is certain to cost more than development of one that has a lower ranking in these two categories. Of course, if the frame is to be used for several surveys or survey rounds, a greater investment in its development can be justified.

Subsections 1, 2, and 3 below discuss in detail each of the three major categories of desirable frame properties. Subsection 4 reviews their interrelationships and summarizes the main points.

1. Quality-related properties

Quality-related properties of a frame are those properties which make it possible to minimize non-sampling errors, especially coverage errors, that might occur because of deficiencies in the frame. The fact that a frame has these desirable properties cannot guarantee that all coverage errors will be avoided, but it does make it easier to avoid them.

The first desirable quality-related property is that the frame consists of <u>well-defined units</u>. The meaning of "well-defined" depends on whether area units or other units, such as housing units and households, are being considered.

Some area units are administrative units: states, provinces, districts, cities, wards, villages, etc., established for various governmental administrative purposes. Higher-level administrative units are <u>usually</u> well-defined in the sense that they have recognized boundaries that are clearly delineated on various types of maps. Their boundaries can usually be recognized or determined fairly accurately in the field, either because they are physical features such as rivers, roads, or coastlines or because persons living near the boundaries know approximately where the boundaries are and know which administrative area they themselves live in.

Lower-level administrative units, such as villages, are not always well-defined in the same sense. Often there are no maps showing the boundaries of these units. It may be possible to establish approximate boundaries by making local inquiries, but this process requires considerable effort and is not always fully successful (see World Fertility Survey, 1975, for further detail on problems that may be encountered in fixing village boundaries).

It is sometimes possible to use area units that are not administrative units. The most common example is the enumeration areas (EAs) established for the most recent census of population. Census EAs may or may not be well-defined. Much can be determined by inspection of

Exhibit 4.2 Checklist of desirable frame properties

- 1. Quality-related properties.
 - a. Well-defined units.
 - b. Adequate identifiers.
 - c. Complete.
 - d. Up-to-date.
 - *e. Stable units.
- 2. Efficiency-related properties.
 - a. Inclusion of accurate and up-to-date supplemental information.
 - *b. Choice of sampling units available.
 - c. Good quality maps of units available.
 - d. Easy to manipulate/process.
- 3. Cost-related properties.
 - a. Low cost of acquisition/preparation.
 - b. Low cost of use.
 - *c. Low cost of maintenance.
 - * Denotes properties that are relevant only or primarily for master sampling frames or other frames to be used for more than one survey or survey round.

the available maps or sketches, but to be safe, the quality of their definition needs to be determined by means of field observation for a carefully selected sample of EAs.

Sometimes small area units are established specifically for use in surveys. In a design where the PSUs are census EAs, for example, the SSUs may be formed, in the sample PSUs only, by using available physical features to divide the EAs into area segments. This process (which is sometimes called "chunking") will be referred to as segmentation in this study. The quality of definition of such areas will depend on how carefully and conscientiously the instructions for forming them are followed. If good maps are available, much of the work can be done in the office. If they are not available, field work will be necessary.

For frames that use non-area units like housing units or households, the first requisite for having well-defined units is that a precise standard definition of the unit be established, i.e., one that covers the wide variety of structures and group living arrangements that may be encountered and explains clearly how they should be treated when preparing listings of such units. The second requirement is that the standard definition be carefully adhered to in practice.

The second desirable quality-related property is that all units have adequate identifiers. As pointed out earlier, a unique identifier is the one item that is indispensable for each frame unit. Usually frame units will have both unique numerical identifiers, referred to here as <u>primary</u> identifiers, and other identifiers, such as names and addresses, which will be referred to as <u>secondary identifiers</u>. The primary identifiers are used for purposes such as manipulating the frame in various ways and selecting samples. They may also be used to link area units in a frame to the maps and sketches, especially when these area units have been established solely for sampling purposes and have no official or generally recognized names. In some countries numerical identifiers assigned to housing units are physically affixed to those units by the use of more-or-less permanent labels.

Particular care is needed in establishing a system of numerical identifiers for units in a master sampling frame. A hierarchical system is desirable, in which the first group of digits identifies the highest-level administrative division in which the frame unit is located, the next group identifies the second-level administrative subdivision, and so on, down to the individual frame units. Some provision for distinguishing urban and rural units may also be helpful.

Secondary identifiers are used primarily to aid in locating frame units on maps and in the fueld. For area units that are administrative subdivisions, the secondary identifiers will be the names of the subdivision and the higher-level administrative units of which it is a part. Names of villages and other small units may not always be unique identifiers, even within higher-level administrative units. In such cases, supplemental information, such as names of roads, rivers or adjacent villages, may be needed. For area units that are not administrative subdivisions, names of the administrative units in which they are located will serve as secondary identifiers.

For housing units, secondary identifiers are names, addresses and identifying characteristics, e.g., "the house with the pharmacy", "the first house on the right beyond the railroad crossing." The inclusion of identifying characteristics in addition to names and addresses is important, especially in areas that do not have clearcut patterns of streets or roads. Listing forms should provide space for recording such information and listers should be trained and motivated to provide it.

For households, the name of the head is the key identifier. The full name should be recorded and its spelling verified if there is any doubt. If households are to be used as USUs, it is important to record the address and other features of the housing units with which they are associated. If there are two or more households in a single housing unit, it may be helpful to identify the members of each household separately on the listing form.

The <u>completeness</u> of a sampling frame has two aspects: the extent to which the intended coverage (as defined in subsection A.3 of this chapter) is actually achieved and the extent to which the desired information for each frame unit (see discussion of content in subsection A.5 of this chapter) is included in the frame.

For frames that cover 100 percent of the target population, completeness of coverage can be checked by cumulating measures of size (such as number of persons or households counted in a census) of the frame units and comparing the totals for administrative areas, such as states, provinces or districts, against control counts for these areas. Maps can sometimes be used in checking for completeness, e.g., to be sure that all of the land area included in each sample PSU has been accounted for by area segments established for use as SSUs.

Different methods are needed to check the completeness of coverage of list frames. There is no obvious way to check the completeness of a list of villages. In many countries there is a ministry responsible for defining and maintaining a complete list of villages and other administrative units. However, the process is sometimes decentralized, and it may be wise to verify centrally-obtained lists by checking with state or provincial officials.

Still other techniques are needed to ensure completeness when the sample design requires that lists of housing units or households for a sample of area segments or villages be used as secondary sampling frames. Whatever the definition of the listing units, the interval between the preparation (or updating) of the lists and their use for sampling and data collection should be as short as possible. Completeness of coverage (and avoidance of the inclusion of units not in the defined areas) will depend in part on the quality of maps or sketches showing the location and boundaries of the sample areas.

Listings of housing units or households are usually prepared by field employees of the organization conducting the surveys. Good quality work, including especially completeness of coverage, can be assured by:

- o Preparation of clear, comprehensive instructions for listing.
- o Training field employees in the use of the forms and instructions.
- o Development of effective quality control procedures.

Two kinds of quality control procedures can be considered. The first, which is appropriate for any listing operation, is for field supervisors to re-list a small sample of the assigned areas, say one or two for each person assigned to the listing operation. If the assigned listing areas are large, the supervisory check might consist of listing only a few (say 10) of the housing units or households in each area to be checked. Whether an area unit is completely or partially re-listed, the checks will be more effective if the re-listing is done without reference to the initial listing and then compared with it.

A second quality control procedure is to compare counts of persons, housing units or households for the areas or villages listed with counts available from another source, such as a population census. Exact agreement of the listing totals with these counts would not be expected; however, large differences in either direction may indicate deficiencies in the listing operation. Tolerances can be established for observed differences between the listing results and the external controls. For each sample area or village failing this check, field supervisors should try to determine the cause of the large difference and, if indicated, arrange to correct deficiencies in the listings.

Frame completeness also requires that specified information items, such as stratification variables and measures of size, be included in the record for each frame unit. If the basic source of frame information (usually a population census) fails to provide this information for some units, actual or estimated values should be developed from other discussed sources. As wi11 be further in connection with efficiency-related properties of frames, measures of size do not necessarily need to be exact.

A frame that is complete at time t_1 will not necessarily be complete at time t_2 . If a frame is to be used only once for sample selection and/or data collection, that use should occur as soon as possible after the frame has been developed. For frames that are to be used more than once, especially master sampling frames which are typically used during a five- or ten-year intercensal period, procedures must be developed for periodic updating to ensure that they are up-to-date.

Changes occur that affect both the number and definition of frame units. Governments frequently create new administrative units or change the boundaries or nature of existing ones. This happens for many Population growth is an important factor: different reasons. some countries have a conscious policy of maintaining upper limits on the population or number of households under the jurisdiction of a single local official. When this limit is exceeded, administrative units are split or restructured in other ways to maintain the desired sizes. As areas become more densely populated, their official status may change from village to town or city and their demographic designation from rural to urban. Less frequently, administrative units may cease to exist, as in Sri Lanka where some entire villages have been inundated as a result of the construction of dams under the Mahaweli Development Programme. Conversely, in this same case new villages have been established in previously unsettled areas because the availability of water for irrigation made cultivation possible.

The smallest administrative units, such as villages and city wards, are likely to change more rapidly than larger ones, such as provinces or states. Table 4.1, which shows changes in the number of administrative divisions and subdivisions in Thailand between 1970 and 1982, illustrates this. Only one new province (changwad) was created during this period, there were substantial increases in the number of smaller but subdivisions: districts, subdistricts and villages. Increasing urbanization led to the creation of new municipal areas and sanitary districts.

	Number of units			Percent increase		
Type of unit	1 97 0	1980	1982	1970-1980	1980-1982	
Province	71	72	72	1.4	0.0	
District First-class Second-class	580 540 40	698 621 77	716 628 88	20.3 15.0 92.5	2.6 1.1 14.3	
Sub-district	5,134	5,930	6,160	15.5	3.9	
Village	45,538	53,387	55,487	17.2	3.9	
Municipal area	120	118	123	-1.7	4.2	
Sanitary distric	t 641	7 0 9	716	10.6	1.0	

Table 4.1 Total number of administrative divisions and subdivisions in Thailand, by type: 1970, 1980 and 1982.

Source: Skunasingha and Jabine, 1983.

Changes that can occur in list frames are well known. New housing units are constructed; existing ones may be demolished deliberately or as the result of natural or man-made disasters. Internal reconstruction or remodelling can increase or decrease the number of housing units in an existing structure. Households change as the result of births, deaths and in- and out-migration.

Use of a sampling frame that is not current can lead to coverage biases. As more time elapses without any action to take account of changes, the size of these biases increases. The use of rules of association to deal with changes in sampling units was discussed earlier, in section B. Other strategies for dealing with the effects of change in using a master sampling frame are discussed in subsection D, 5 of this chapter. Strategies for dealing with change in other types of frames are discussed in section E.

If there is a choice with respect to the kinds of units to be used in a frame, it is preferable, other things being equal, to choose the most stable kinds of units available, i.e., those that are least subject to change in number, definition and size. Procedures for dealing with changes in frame units tend to be costly and complex; therefore, minimizing the need to use such procedures should be a major objective in frame development.

All types of units commonly used in frames are subject to change in some degree. This is even true of areas defined entirely by physically identifiable boundaries: the course of a stream can change; old roads can be torn up. However, priority can be given to the kinds of units that are least likely to change over time. Administrative units with defined boundaries should be preferred to administrative units whose boundaries have not been clearly established, e.g., villages in many countries. In some cases, area segments with physically identifiable boundaries might be preferred to administrative units, some of whose boundaries are imaginary, i.e., do not correspond to physical features such as roads, railroads and bodies of water. However, such area segments should not cross boundaries of administrative units for which separate survey estimates are needed; therefore, some use of imaginary boundaries is virtually inevitable. Finally, it should be clear that housing units are more stable than households; both change, but households are likely to change at a substantially higher rate.

2. Efficiency-related properties

Efficiency-related properties of frames are those qualities that make possible and facilitate the use of efficient survey designs. Efficiency in this context refers to the relationship between sampling error and the cost of producing survey estimates; the most efficient survey design is the one that produces the desired level of precision at the lowest possible cost. Perhaps the most important of these properties is the <u>inclusion in</u> the frame of accurate, current supplemental data for each frame unit. Measures of size, such as population, number of households and number of operators of agricultural holdings, are especially useful. Measures of size can be used in the following ways:

> o To construct sampling units for which the listing or interviewing workloads can be predicted with reasonable accuracy.

o To form strata of units classified by size.

USES FOR

MEASURES OF

o To determine the allocation of sample PSUs to SIZE OF FRAME strata.

- UNITS o To select units with probability proportionate to size (PPS).
 - o As auxiliary variables for ratio or regression estimates.

The availability of measures of size is especially important if there is large variation in the size of frame units. If there are no measures of size, units must be selected with equal probability. With measures of size, any of the more efficient sample design and estimation techniques listed above can be used.

Nearly all of the master sampling frames described in the case-studies contain one or more measures of size, the commonest being counts of population, households or housing units. If frames are to be agricultural activities, of used for measures size related to agriculture are sometimes included. The frame for the U.S. master sample of agriculture (see Case-study: United States of America) included the indicated number of farms (as determined from county highway maps) for each count unit.

The master sampling frame developed from Nigeria's 1973 population census did not include measures of size, because the census results had been discarded. Consequently, a two-phase master sample design was used. For most of the primary strata, the sample for phase one consisted of census enumeration areas (EAs) selected with equal probability. Housing unit listings were prepared for the sample EAs, following which, subsamples of EAs were selected with probability proportionate to total number of households in urban strata and proportionate to number of farm households in rural strata. A more efficient sample design could have been developed had the census counts been available.

Measures of size are sometimes included in housing unit or household listings used for the final stage of sampling in multi-stage designs. One can expect a significant correlation between the size of a household (number of persons) and its income and expenditures. For this reason, it is a fairly common practice, in income and expenditure surveys, to order or renumber the households in a listing unit by size and to select the indicated number or fraction of households systematically with a random start.

Supplemental information for frame units can also serve as a basis for stratification of frame units. Indeed, some such information is virtually essential in a master sampling frame. Three kinds of information are often used:

- (1) Identification of the major administrative units, such as states, provinces, districts, metropolitan areas and cities, in which the frame units are located.
- (2) Information about the urban or rural character of the frame units.
- (3) Information about population density.

Some examples of the second category, i.e., attributes describing the urban or rural character of frame units, taken from the master sampling frames described in the case-studies, are listed in Exhibit 4.3.

Measures of population density can be used to establish strata, for each of which the optimum number of stages of sampling and ultimate cluster sizes can be determined separately. An example of this is provided by the case-study for Australia. Other attributes, such as principal economic activities, predominant ethnic groups and median income, to name but a few, can also be useful for stratification.

Errors in supplemental data do not necessarily lead to biases in survey results. They can, however, limit the efficiencies that can be gained from using the supplemental data for sample design or estimation Therefore, reasonable efforts to make the supplemental data as accurate as possible are worthwhile.

A desirable property for a master sampling frame is that it be constructed so that a <u>choice of sampling units</u> is available for surveys that differ in their data requirements. While relatively small areas, such as census enumeration areas, may be the best choice for PSUs in some surveys, cost considerations may lead to the choice of larger PSUs in other surveys. These larger PSUs might be formed from groups of adjacent enumeration areas or from administrative units, such as urban wards and rural districts. National surveys tend to use larger PSUs than surveys designed to produce subnational estimates, and surveys of relatively rare populations usually require larger PSUs than those designed to cover the general population (for a more detailed discussion of the choice of PSUs in multi-stage samples, see Kish, 1965, Chapter 10).

Countries	Attributes
Botswana	Towns Villages Lands Freehold farms Cattle posts
Могоссо	Urban Luxurious Modern New Medina Old Medina Industrial area Shantytown Urban <u>douar</u> Small urban centre Rural
Nigeria	State capitals Large towns Other towns Rural
Saudi Arabia	Metropolitan Other urban Rural
Sri Lanka	Urban Rural Estate (large holdings)
Thailand	Municipal areas Villages In sanitary districts Other
United States ¹	Incorporated places Densely populated unincorporated areas Open country

Exhibit 4.3 Attributes used to classify frame units by urban-rural characteristics in selected countries

.

¹ Frame for master sample of agriculture.

There are two things that can be done in the development of a master sampling frame to facilitate flexibility in the choice of sampling units. The first is to organize the frame units in a <u>hierarchical</u> <u>structure</u>. This can be illustrated by the frame developed from Sri Lanka's 1981 census of population. The smallest (elementary) frame units are census blocks, most of which have fewer than 100 housing units. The blocks are identified and listed according to the following hierarchical structure:

Districts

Municipal, urban, or town council areas (Urban) Wards Census blocks

(Rural) Assistant government agent (AGA) divisions (Rural) Grama sevaka (GS) divisions Census blocks

Thus, each district is made up of municipal, urban and town council areas and AGA divisions. Each municipal, urban or town council area is divided into wards and each ward has one or more census blocks. In the rural areas, AGA divisions are divided into GS divisions and the latter are divided into census blocks.

Recent household surveys in Sri Lanka, which are designed to produce sample estimates by district, use census blocks as PSUs, combining adjacent blocks where necessary to have at least 20 housing units in a PSU. However, a household survey designed to produce only national estimates might use larger units as PSUs, e.g., wards and GS divisions or larger groups of adjacent census blocks. Note also that a frame with this structure could be used for more than one stage of sampling. For example it would be possible to select a first-stage sample of wards and GS divisions and a second-stage sample of census blocks within the sample wards and GS divisions.

The second thing that can be done to enhance flexibility in the choice of sampling units is to assign identifiers to frame units (both the elementary and higher-level units) on the basis of geographic contiguity. A serpentine arrangement is generally considered appropriate for this purpose, as shown in Exhibit 4.4.

Exhibit 4.4 Serpentine numbering of elementary frame units within the next higher-level unit.



The same technique may be used, of course, to number units at each level within the units at the next higher level.

This kind of numbering would make it theoretically possible, if the frame were computerized, to write a program that would form PSUs by grouping elementary frame units according to an appropriate set of rules. However, a much better job could be done with the help of maps showing the locations of the elementary units. In Exhibit 4.4, for example, combining units 01, 02 and 06 might be preferable to combining units 01, 02, and 03.

A second advantage of this kind of identification system is that it makes possible the selection of a geographically dispersed sample of PSUs or SSUs through the use of systematic sampling

The importance of using well-defined frame units has already been pointed out in subsection C,1, which dealt with quality-related properties of frames. For area units, good definition depends to a large extent on the availablility of accurate, detailed, large-scale maps showing the boundaries of each unit.

The availability of such maps can also contribute to the efficiency of sample designs. When costs are taken into account, a design that uses small (say 5 to 15 housing units) compact area segments is likely to be more efficient than one that uses larger segments with listing and subsampling of housing units. However, if the maps needed to define the small area segments are not already available, it may be preferable to use the larger segments with listing and subsampling. An alternative is to prepare maps of the desired quality only for a sample of PSUs or SSUs. This option discussed further in connection with master samples in Chapter 5. Another frame property that facilitates the use of efficient sample designs is <u>ease of manipulation</u> of the records for the frame units. The obvious way to achieve this property, at least for master sampling frames (which often contain several thousand elementary units), is to computerize the frame. Techniques such as sampling with probability proportionate to size can be used much more readily with a computerized frame, and some techniques, such as cluster analysis to form strata or controlled selection, would be virtually impossible to use if the frame were not computerized. The gains from using these more efficient sampling techniques must be balanced against the costs and time needed to develop a computerized frame. The costs of computerizing a frame can, of course, be more readily justified for a master sampling frame than for a frame that is to be used only for a single survey.

Ease of manipulation is also desirable for frames from which sampling is to be done manually, the primary example being listings of housing units or households for sample PSUs or SSUs. Listing forms should be designed to facilitate sample selection. Obviously, this can only be done if the specific sample selection process to be used has been decided on prior to the design of the listing form. Two examples follow:

<u>Example 1</u>: The sample design calls for the listing of housing units, followed by selection of a systematic sample of those housing units operating one or more household economic activities. The listing form will, of course, require a column to indicate the presence or absence of such activities in each housing unit. In addition, it would be desirable to include an adjacent column in which consecutive serial numbers could be assigned to housing units with economic activities, so that the designated systematic sampling pattern could be applied directly to those serial numbers.

Example 2: For an income and expenditure survey, the sample design calls for selection, from segment listing forms, of a systematic sample of households, with the households in each segment ordered by number of persons. Households will be listed in the order in which they are visited, and the number of persons in each recorded in a designated column of the listing form. An adjacent column, marked "Office use" should be reserved for sample selection. Sample selection could proceed as follows:

- (a) In the office use column, assign a new consecutive series of household serial numbers, starting with all of the one-person households, proceeding to the two-person households, and so on.
- (b) Apply the designated sampling instructions to determine the serial numbers of the households to be selected, e. g., 8, 23, 38, 53, etc.

(c) locate and circle each of the designated serial numbers in the office use column.

There are, of course, many acceptable methods of sampling from listing forms. The important point is to decide in advance what sample selection method to use and then to design the listing form in a way that will facilitate the selection process and minimize errors of implementation.

3. Cost-related properties

Properties that favour quality and efficiency in the use of sampling frames usually have costs attached to them; sometimes the costs are substantial. If two alternative frame sources would result in the same quality and efficiency, the one with the lower costs of development, use and maintenance would obviously be preferred. Decisions about frames are seldom that simple, however; a careful weighing of costs and benefits is usually needed. Furthermore different types of costs associated with frames are interrelated. Resources invested in initial development of a master sampling frame can reduce subsequent costs of its use and maintenance or updating.

The development costs of frames, be they master sampling frames or secondary frames, depend largely on what resources are already available. The resources most needed are:

- o Lists of well-defined administrative units or units established for census enumeration purposes.
- Measures of size for these units that are reasonably current and accurate.
- Maps that (1) show the boundaries of the administrative units and (2) have sufficient detail so that the administrative units can be subdivided into smaller area units.
- o (For secondary sampling frames only) current lists of structures, housing units or households.

It will be an added advantage if the lists of administrative units and their measures of size are already computerized.

The resources needed to develop frames for surveys are essentially the same as those needed to develop frames for conducting censuses of population and housing. Low costs of frame development can best be achieved, therefore, by treating the development, maintenance and updating of frames for censuses and household surveys as a single, integrated ongoing process. Frame development costs which would appear excessive for a single survey can be justified much more readily as part of an integrated programme of censuses and household surveys. The costs of developing accurate, detailed maps can be substantial, even in the context of an integrated programme of censuses and surveys. If such maps have not been developed already for non-statistical purposes, it may be necessary, at least in the short run, to rely on alternatives. This explains why some countries rely on list frames of administrative units, especially for rural areas. Examples are Ethiopia, which uses a list of farmers' associations, and Thailand, which uses a list of villages. Another alternative, which will be discussed in connection with master samples in Chapter 5, is to develop detailed maps only for a master sample of higher-level administrative units.

Development costs are also a consideration in deciding what kinds of secondary sampling frames to use. The principal alternatives are listing housing units or households vs. segmentation, i.e., dividing each sample PSU or SSU into compact area segments in which all households will be interviewed. Sometimes a combination of these methods may be used. If reasonably detailed maps are available, the cost of segmentation is likely to be less than that of listing.

The costs of using a master sampling frame are determined primarily by the medium on which it is stored. Given the existence of reasonable automated data processing capabilities, the costs of selecting samples should be considerably smaller if the frame is computerized. With respect to frames for sampling within PSUs or SSUs, there is probably no appreciable difference in cost between sampling from listings and selecting a sample of compact segments from a set of PSUs or SSUs that have already been subdivided into such segments.

The costs of <u>maintaining</u> master sampling frames depend primarily on the stability of the frame units. As pointed out earlier, the most stable units are those defined entirely by physically identifiable boundaries; such units seldom require redefinition, whereas in most countries new administrative units, especially at the lower levels, are created from time to time in response to changes in population size and distribution or for other reasons. Either of these types of frame units is, of course, subject to changes in size (population or number of households). Strategies for dealing with such changes are discussed in subsection D; the costs should be similar for the two kinds of frame units.

Many IHSP designs call for use of the same sample of PSUs or USUs in more than one survey or survey round conducted at different times. For such designs, the cost of maintaining the secondary sampling frames needed for these units is an important consideration. A frame consisting of a set of small area segments defined in each sample PSU or USU is clearly easier to maintain than a frame that is a listing of housing units or households. If the choice of units lies between housing units and households, the former, being more stable and easier to identify, are preferable in most circumstances. In some countries, the cost of updating a housing unit listing can be reduced by physically labeling each unit with a unique identifier during the initial listing.

4. Summary

This section has identified desirable properties of frames and has explained how key decisions, especially as to the choice of frame units and storage medium, can influence the extent to which a frame will have those desirable properties. The necessity of evaluating tradeoffs between quality and efficiency, on the one hand, and cost, on the other, has been pointed out. The quality and efficiency that can be built into a frame will depend, to a large extent, on the resources already developed for other purposes (including population censuses) that are available for frame construction and on the number and timing of surveys for which the frame is to be used.

The checklist of desirable frame properties presented at the start of this section (Exhibit 4.2) can be used for the evaluation of existing frames, to determine where improvements are possible. It can also be used to evaluate alternate proposals for the construction of new frames.

The treatment of frames in this section has been broad in scope, covering both master sampling frames and frames used for sampling at the later stages of multi-stage sample designs. The next section of this chapter will cover the development of a master sample frame <u>as a</u> <u>process</u>, providing a systematic account of the steps required. The final section will explore procedures for the development of secondary sampling frames, i.e., frames that are needed only for a sample of PSUs or USUs and are not available directly from the master sampling frame.

D. <u>Development and maintenance of a</u> master sampling frame

Preceding sections of this chapter have described the attributes and desirable characteristics of all types of frames used in household surveys. This section covers the process of developing and maintaining one particular type of frame: a master sampling frame (MSF).

As explained in Chapter II, D, an MSF is one which covers the entire survey population and is intended for use to select samples for several different surveys or for different rounds of a continuing or periodic survey. The MSF has a key role in an integrated programme of household surveys (IHSP): use of an MSF for more than one survey or survey round is a significant form of integration and can lead to important savings in the cost per survey of frame construction.

Exhibit 4.5 lists the main steps in planning the development of an MSF. In reality, planning does not always follow an orderly sequence like the one shown: findings at later stages or changes in requirements

may lead to changes in earlier decisions. However, the five steps usually proceed more or less in the sequence in which they are shown and will be discussed in that order in subsections 1 to 5 below.

Exhibit 4.5 Steps in planning the development of a master sampling frame

- 1. Determine general objectives and strategy.
- 2. Identify and evaluate available inputs.
- 3. Decide on key characteristics.
 - a. Coverage
 - b. Frame units
 - c. Record content
 - d. Storage and processing medium
 - e. Auxiliary materials
- 4. Prepare schedule for initial development operations.
- 5. Develop plan for updating.

The timing of MSF development in relation to the overall country programme of population censuses and household surveys is critical. Aб recorded in the case-studies, most of the MSFs developed for household surveys rely entirely or largely on population censuses for inputs. The suitability and quality of such inputs will be served best by doing much of the initial planning for MSF development just prior to the census enumeration. The census itself requires a frame similar to that which will be needed for household surveys. Census products, such as population or household counts and enumeration sketch maps for small areas, can be used to enhance the quality of the census frame for sampling purposes. In fact, another way of describing the ideal process for MSF development would be as follows: (1) develop the frame needed for the population census, (2) enhance the census frame with census outputs and (3) structure these materials in a form suitable for the anticipated sample selection operation. To follow this ideal sequence, one must start with a reasonably clear picture of the purposes for which MSF will be used. Once these requirements are established, the steps necessary to meet them can be incorporated in the census plan.

This process can also work in the opposite direction. If an MSF is developed from a census and is updated appropriately during the subsequent 5- or 10-year period, the job of developing the frame for the next census is likely to be considerably easier. These relationships between the planning and operational stages of the population census and the MSF are shown graphically in Exhibit 4.6. In this depiction, it is assumed that the MSF is initially planned and constructed during the first census period and that it is fully updated during the second census period. Thereafter, the cycle may continue indefinitely.

Naturally it will not always be feasible to follow this ideal process. The need for an MSF may become evident midway between two censuses, at which point some of the census inputs will be at best outdated and at worst unavailable. Further, there are a few countries that have not conducted population censuses but may nevertheless decide to undertake a programme of household surveys. In such situations, compromises may be necessary, and desirable improvements in the quality of measures of size and maps may have to await the next (or first) population census.

Nevertheless, a review of the case-studies shows that most of the MSFs were developed primarily from census materials within a period of not more than two or three years following the census enumeration. In a subset of these cases, features were built into the census plan specifically for the purpose of facilitating the construction of an MSF. In Thailand, for example, all census enumeration areas in municipal areas were subdivided into blocks, and population counts were obtained by block so that the blocks could be used as the primary frame unit for post-censal household surveys.

In this section, therefore, the discussion will focus primarily on the common practice of constructing an MSF from the population census, in conjunction with the somewhat less common (but desirable) feature of considering the MSF requirements as part of the census planning process. Subsections 1 to 5 will cover the five major steps in the development process, as shown in Exhibit 4.5. Subsection 6 takes a look at some alternatives when there are no inputs available from a recent census of population.

1. Determination of general objectives and strategy

One should begin the development process by asking: For what purposes will the MSF be used? The obvious answer is: to select samples of first-stage and in some instances second-stage units for one or more household surveys or survey rounds.

Keeping in mind the benefits of integration of statistical programmes, however, planners should consider whether the MSF might also serve other purposes. For example, could the MSF be developed in conjunction with a hierarchical listing of the country's administrative subdivisions which, with periodic updating, could provide the geographical framework for all types of censuses and surveys conducted by the statistical office? Going a step further, could there be a tie-in of some kind between the MSF and a statistical data base containing aggregate data, from a variety of sources, for different kinds of administrative divisions and subdivisions? The aggregate data,

Exhibit 4.6 The population census and the master sampling frame (MSF)

Populati	lon census	Mas	ter sampling frame
FIRST	Preparation	~~~ ~~ ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	Plan for MSF
CENSUS	Enumeration		
1211202	Processing		Construct MSF
INTERCEN	ISAL PERIOD		Use MSF*
SECOND	Preparation		Plan for full update of MSF ^{**}
CENSUS	Enumeration		
	Processing		Full update of MSF
INTERCEN	SAL PERIOD		Use updated MSF*

- * Some updating is likely to be necessary during intercensal periods, but full updates should be tied to censuses.
- ** The existing MSF would continue to be used as needed until updated during the processing phase of the second census.

which would be of interest in their own right, could also be used as measures of size or to construct stratification variables for sampling purposes.

To plan and develop such an integrated multi-purpose small-area data system requires a certain level of sophistication in systems development and may not, therefore, be within the reach of all countries as they initiate IHSPs . However, the potential benefits of such an integrated system are sufficient to make it well worth considering as a longer-term goal.

Getting back to the basic purpose of the MSF, the next question might be: How many different samples of PSUs are likely to be needed? It is unlikely that this question can be answered definitively prior to the population census. It is possible that a single master sample of PSUs could meet all requirements for household surveys during the intercensal period, especially if the country opts for the first class of IHSP design described in Chapter III., Section D, i.e., a single multi-subject continuing or periodic survey. However, even if this is the initial choice of design, unanticipated needs may arise during the period for which the MSF is expected to be used. For example, future changes in the structure, size or basing arrangements of the field staff for interviewing might make it desirable to shift to a sample design with smaller PSUs, e.g., census enumeration areas, in place of some larger administrative subdivision. Conversely, special surveys to measure relatively rare phenomena, such as disability, might require PSUs or SSUs larger than those deemed appropriate for the basic multi-subject survey.

Possibilities like these favour the development of an MSF with flexibility to meet varying requirements, provided the cost is not substantially more than that of an MSF designed solely to meet known requirements. A model which offers this kind of flexibility and has been used by many countries is the MSF which uses census enumeration areas (EAs) as the basic frame unit and organizes them hierarchically and geographically so that the frame can be used in any of the following ways:

- o To select samples of EAs directly.
- o To form PSUs of any desired size by grouping adjacent EAs.
- o To select samples of administrative subdivisions and, if desired, to sample in one or more additional stages down to the EA level.

This frame structure is, indeed, essentially the same as that needed to conduct the census. With proper advance planning, this kind of MSF can be produced at a very reasonable cost as one of the products of census processing. Even greater flexibility will be assured if the MSF is produced in computer-readable form.

Some countries have developed MSFs that have frame units smaller than census EAs. In Thailand, for example 1980 census EAs in municipal areas (the larger urban places) were subdivided, using physical boundaries, into blocks, and census counts of population and households were compiled by block. These blocks were used as the basic frame units in the municipal area component of the MSF, and blocks or groups of blocks are used as PSUs in most household surveys. It was possible to do all of this at a reasonable cost because good quality street maps were already available for virtually all of the municipal areas.

This illustration suggests two other questions that should be considered in the pre-census planning for development of an MSF. The first is: <u>Should the basic structure of the frame be the same for all</u> <u>kinds of areas?</u> In Thailand (see case-study), two quite different structures were developed -- one for the municipal areas and one for all other areas -- primarily because of the differences in the quality of maps available for the two sectors. However, the majority of countries covered by the case-studies use a single frame structure for all geographic sectors.

The second question is: What procedures should be made part of the census in order to enhance the quality of the MSF? The answer to this question depends on the cost of the proposed procedures and on the extent to which they will serve the purposes of the census itself.

Consider, for example, the possibility of requiring each census enumerator to prepare a sketch map of his or her EA. The precise requirements for doing this might vary depending on how the EAs were defined initially, e.g., by maps showing their boundaries, by a written description of the boundaries, or simply by the name of a village or other locality. Typically, the enumerator would be asked to prepare a sketch showing: the EA boundaries; the streets, roads and other physical features within the EA; and the location of dwelling units and other structures listed or enumerated.

Such sketch maps would contribute to the quality of coverage in the census itself. They would help census enumerators to do a more systematic job of canvassing their EAs and they would be useful for supervisory and other checks on the work of enumerators. With one additional step, the sketch maps could serve as an important adjunct to the MSF. That step is simple enough in theory: make sure that all sketch maps are retained after the census enumeration and brought together in a central location.

Another possible use of the sketch maps may have occurred to the reader. Why not use them to divide EAs into smaller areas for sampling purposes? Here is where the costs must be carefully considered. A

review of the census sketch maps and the associated listing forms might show that for 60 percent of the EAs the segmentation could be carried out according to specifications in an office clerical operation; for the remaining 40 percent, some kind of field work would be necessary. Significant costs would be associated with both the office and field operations, especially the latter. It would not be economical, then, to do the segmentation for all EAs, since the results would be used only for EAs selected for inclusion in one or more samples. It is even debatable to what extent it would be worthwhile to attempt, through more intensive enumerator training and supervision, to improve the quality of the sketch maps prepared in the census, since only a small proportion of them would be used for sampling purposes.

The main points presented in this subsection can be summarized as follows:

MSF	OBJECTIVES	0	To the extent possible, determine the number and broad character of the different samples of PSUs that will be needed during the period for which the MSF will be used.
		0	Consider making the MSF part of an integrated, multi-purpose small-area data system.
MSF	STRATEGIES	0	Build in flexibility. If possible, use census EAs as basic frame units in a hierarchical, geographically-ordered structure.
		0	Adopt census procedures that will benefit the MSF, provided that they are needed for census purposes or have low marginal costs.

2. Identification and evaluation of available inputs

Ideally, initial planning for development of an MSF should coincide with the preparatory phase of a population census, with the understanding that the census frame, enhanced by data and other census outputs, will be the principal input to the MSF. Even in this ideal case, however, it would be useful to inventory and evaluate all potential sources of inputs to the MSF. Sources other than the population census may provide useful supplemental data and maps for the frame units. More significantly, most MSFs require some updating during intercensal periods; other sources must be relied on for this purpose.

If the MSF is to be developed with little or no access to usable census materials, a complete inventory of potential sources is, of course, essential. Even in the best of circumstances, the development of an MSF for household surveys is likely to have significant costs; therefore existing materials should be used as much as possible. In broad terms the inventory should cover three types of materials:

		0	Lists of administrative defined areas.	subdivisions	and	other
WHAT LOOK	TO FOR	0	Data, e.g. population or counts, for these areas.	h ousehol d		
		о	Maps			

These materials can be found in many different places. The obvious place to start is within the national statistical organization or system, but there are many other possible sources that should be investigated.

The inventory within the national statistical organization or system can be approached in two ways: by organization and by programme. Many statistical systems have a central geographic/cartographic unit: this is the obvious place to begin the search for lists and maps. The programmatic approach looks systematically at recent and planned censuses and surveys. Priority would be given to an upcoming or recent population census. Other programmes that might provide inputs to an MSF include agricultural censuses and surveys and surveys of administrative subdivisions. An example of the latter is the programme of village surveys conducted by the National Statistical Office of Thailand. Each year all village headmen are asked to complete questionnaires giving number of persons, principal economic activities, area cultivated in major crops and several other items for their villages. Potential uses of these kinds of data in updating an MSF are evident.

Looking outside of the statistical system, an initial question might be: what ministry or department has the primary authority and responsibility for internal administration of the country and, more specifically, for establishing and updating the structure of the country's administrative subdivisions? Commonly, this function will reside in some unit of the interior ministry or its equivalent.

A second thing to look for is a unit of the government that provides centralized cartographic services. These services include both the original production of topographically accurate maps drawn to scale and the updating of such maps, taking account especially of changes in administrative subdivisions and place names. In some countries, such units also have access to aerial photographs or other remote-sensing imagery, which may be useful at the later stages of sampling.

In addition to these two primary sources of frame inputs from outside of the statistical system, there are several other places to look. Any ministry or department that has programmes with field operations serving all or a large part of the country should be contacted as a potential source of lists, data and maps. Such widespread activities are perhaps most likely to be found in the areas of health, housing and agriculture. If the country has an active public housing programme, it will be important, for updating the MSF, to know the locations of new housing projects.

In many countries, it would be desirable for the inventory of potential sources to cover both units of the central government and their counterparts at the next lower level, e.g., states, provinces or districts. The latter may sometimes have more detailed or current information; on the other hand, the nature and quality of their inputs may be quite variable. If the statistical office has a field staff in each of these administrative divisions, it should be much easier to locate, evaluate and use their materials.

The next step, after identifying potential inputs to the MSF, is to evaluate them. What criteria and methods of evaluation should be used? For evaluation criteria, the reader may turn back to Exhibit 4.2, which appears earlier in this chapter, at the beginning of Section C, Desirable properties of frames. From that list of desirable properties, one can pick certain ones that are of major importance for each of the three kinds of frame inputs: lists, data and maps.

For lists, the first set of criteria has to do the with characteristics of the units that are to be considered as potential frame units. What kinds of units are they, how well defined are they, what kinds of identifiers do they have, how stable are they, and how do they relate to administrative divisions and subdivisions for which separate survey estimates may be needed? A second aspect is the coverage of the list. Does it cover the entire country or does it omit some areas, deliberately or otherwise? If the latter, what is the nature of the omissions? Third, how current is the list now and how often and when is it likely to be updated? Finally, in what forms can it be made available for use in developing the MSF and at what costs?

In evaluating potential sources of supplemental data for the frame units (i.e., measures of size and stratification variables), the first question is: for what potential frame units are the data items available? Second, how accurate are the data items? Accuracy in this context refers to differences between the data items and their "true values" for the reference date or period to which they refer. Third, how current are the data and how often (if at all) and when are they likely to be updated? Finally, in what forms are the data available (e.g., publication, computer printout, file cards, computer tapes, etc.) and what are the probable costs of obtaining and processing them for use in the MSF?

In the evaluation of maps, the first question, as in the case of data, relates to potential lists and frame units. For what administrative and other units are the locations and boundaries shown? Second, what is the scale of the maps and how much detail is shown in terms of natural and man-made physical features? In particular, is there enough detail so that the maps could be used as a basis for segmentation of administrative subdivisions or census EAs? The answers to these questions might be different for maps of large cities and for maps covering smaller places and rural areas. Third, when were the maps prepared and is there any regular system for updating them? Finally, what facilities are available for reproduction of the maps and what are the probable costs of acquisition and reproduction, if needed?

Given these evaluation criteria, what specific procedures should be used to conduct the evaluation of potential inputs to the MSF? The evaluation process can logically proceed in three stages: review of the specifications and procedures used to create the materials and to update them; systematic inspection of the materials; and external validation. For materials to be developed in an upcoming census of population, the second and third stages of evaluation can only be performed after the census. However, the pre-census evaluation of the procedures for producing the frame inputs should pay special attention to the need for built-in checks and other features to assure the quality of the results.

The need for a careful review of available documentation of the specifications and procedures for creating the lists, data and maps to be used as frame inputs should be obvious. Such documentation may include instructions to field and clerical personnel, questionnaires, forms, record formats, computer edit specifications, and a variety of other materials. If appropriate documentation is not available, this can be taken as an indication of potential quality problems and the need for special care in the inspection and validation phases of the evaluation.

Procedures used to inspect materials will vary according to the materials. For lists, it may be useful to count the entries and compare with control totals, if available. The system of identifiers should be carefully reviewed. If possible, statistics should be developed on the frequency of additions, deletions and other changes to the lists. In some cases it may be possible to compare lists with maps to check the lists for completeness.

Standard edit procedures can be adopted for internal checking of the quality of data -- measures of size and stratification variables -- for lists of potential frame units. The frequency of missing data for each variable should be determined. Range checks can be performed for quantitative items, and qualitative items can be checked for invalid entries or codes. Such edits will, of course, be more readily performed if the files have been computerized. In the inspection of maps, the major elements to look for have already been stated, i.e., the extent to which the boundaries of various administrative subdivisions and other units are clearly shown, and the amount of physical detail available for segmentation of administrative subdivisions. If the maps have been produced by a trained cartographic unit, the quality is likely to be uniform from one area to another. On the other hand, a review of sketch maps produced by census enumerators would probably find considerable variation in quality. An effort should be made to estimate the proportion of sketch maps that meet clearly defined standards of quality.

Procedures for external validation of potential MSF inputs depend on what kinds of related materials are available. The ideal procedure for validating maps is to check them in the field, to the extent that resources are available to do this. If field checks are made, the opportunity should be taken to make corrections and add details. Census population or household counts at various levels of aggregation might be checked against counts or estimates from other sources: a civil register, a census of agriculture, health department records, etc.

When dealing with large lists, data files and sets of maps, sampling can be a useful tool for internal and external validations. For example, to evaluate the quality of sketch maps produced in a census, one might select a relatively large sample for clerical inspection and then select subsample of these for field review. Likewise, the inspection of a computer printout of census data by enumeration area might be done by checking a systematic sample of lines, i.e., every nth line.

Inventory and evaluation of potential inputs is an important step in the development of an MSF. For convenience, the main points discussed in this subsection are summarized in Exhibit 4.7.

3. Decisions on key characteristics

After the general objectives of the MSF have been determined and the potential inputs inventoried and evaluated, the next step is to make decisions on the key frame characteristics. These are:

- o Coverage
- o Frame units

KEY CHARACTERISTICS OF MSF

- o Record content
- o Medium
- o Auxiliary materials

Exhibit 4.7	Summary: inventory and evaluation
	of potential MSF inputs

WHY AN INVENTORY?	Even if population census is to be main source, other sources needed for supplementary data and updates.
WHAT TO LOOK FOR	Lists, data, maps.
WHERE TO LOOK	<pre>Statistical office Geographic/cartographic unit Programmes: population, housing and agriculture censuses, surveys of local governmental units, etc. Other national government units Ministry of interior/home affairs Cartographic service Other ministries, e.g., health, housing, agriculture Sub-national government units</pre>
KEY EVALUATION CRITERIA	Lists Characteristics of units Coverage Current validity Medium Data Units for which data are available Accuracy Current validity Medium Maps Details available for subdividing units Current validity Format and reproducibility
EVALUATION PROCEDURES	Review of documentation Inspection of materials* External validation*

*Use sampling when appropriate

Once these key decisions are made, detailed planning of development operations can proceed.

The decisions should be informed by careful consideration of the desirable properties of frames, discussed earlier in Section B of this chapter; and by a recognition that tradeoffs among quality, efficiency and cost will be necessary. It will also be necessary, at this stage, to have at least a rough idea of the sample designs that will be used in the surveys to be conducted during the MSF's scheduled period of use. In particular, answers to the following sample design questions are needed:

- o Will the PSUs be formed from administrative subdivisions, census EAs, or area segments within census EAs?
- o What is the minimum size (in terms of population, housing units, etc.) of PSU that will be acceptable?

These considerations will, in turn, determine whether it will be feasible to select, from the MSF, a master sample that can be used to satisfy the sample requirements of all or some of the planned surveys. If there is to be a master sample, its requirements should be given priority in deciding on the key characteristics of the MSF.

The cost of developing the MSF should be controlled by <u>not</u> including features that are needed only for sample PSUs, unless such features can be incorporated at little or no cost. For example, suppose it has been decided that a master sample of census EAs will be selected and that sample EAs will be subdivided into area segments to be used as SSUs. Furthermore, it is expected that field work will be necessary in a significant proportion of the EAs to provide the basis for the subdivision or segmentation. In this situation, it would clearly be wasteful to do the field-work for all of the census EAs.

It may also be useful, in approaching these key decisions, to be aware of features that other countries have incorporated in their MSFs. Exhibit 4.8,. which is based on the country case-studies in Appendix A, shows the coverage properties, basic frame units, and measures of size (part of the record content) in the MSFs developed by the 11 countries. Existing practice is, of course, not always the best guide for decisionmaking, but for these particular MSF features, it would appear to provide useful guidelines.

Decisions about <u>coverage</u> are relatively straightforward. Most countries opt for full national coverage. For the time being, Ethiopia's household surveys cover only the rural areas, so the urban sector is not included in the MSF. Some areas or groups are excluded because they have special characteristics or are difficult or costly to

Country (1)	Coverage (2)	Basic Frame Units (3)	Measures of Size (4)
Australia	1. National, regular dwellings 2. National, special dwellings	Census EAs Special dwellings	No. of DUs No. of occupants
Botswana	l. National, urban 2. National, rural	Census EAs Blocks within EAs	No. of HHs No. of DUs
Ethiopia	Rural only, some areas excluded	Farmers' associations	Estimated no. of members
India	l. National, urban 2. National, rural	Census blocks Census villages	Population Population
Jordan	National	Census EAs	No. of HHs, population
Morocco	 National, urban National, rural, excluding sparse areas 	. Census EAs Communes	No. of HH s, population No. of HH s, population
Nigeria	National	Census EAs	None
Saud1 Arabia	National	Emirates	Population
Sri Lanka	National	Census EAs	No. of HHs, population
Thailand	 National, municipal areas 	Blocks within census FAs	No. of HHs, population
	 National, other areas, excluding hill tribes 	Villages	No. of HHs, population
United States	Open-country areas	Count units (area segments)	No. of farms, no. of DUs

Exhibit 4.8 Key characteristics of master sampling frames for selected countries

Source: Case-studies Col. (4) symbols: DU-dwelling unit, HH-household, HU-housing unit

interview in household surveys. Ethiopia's MSF excludes 2 of the country's 14 regions entirely and selected parts of the remaining 12 regions. Nomadic groups, estimated at 5 percent of the population, are also excluded. Morocco excludes sparsely populated areas, containing about 10 percent of the rural population, and Thailand excludes its hill tribes. Note that where such groups or areas are excluded from the survey populations for particular surveys, it still may be desirable to include them in the MSF if relevant information is readily available from a census or other source, in order to provide flexibility in making coverage decisions for future surveys.

The target population for a household survey can be divided into three categories according to their places of residence: regular dwelling places, special dwelling places and institutions. In theory, if the frame units and PSUs are well-defined areas, all three groups can be covered by sampling these units. However, if persons in special dwelling places and institutions are to be included in the survey population for some surveys, it may be more efficient, from a sampling point of view, to maintain a separate list frame covering one or both of these two groups. An example of this practice is found in the case-study for Australia.

Finally, with respect to coverage, in some countries the basic features of the MSF differ substantially for urban and rural areas, because of basic differences in the available frame inputs or the anticipated sample designs for the two sectors. Examples can be found in the case-studies for Botswana, Morocco and Thailand. It will be convenient in such cases to refer to the urban and rural frame components of the MSF.

The most important of the key decisions on frame characteristics is the choice of the kinds of frame units to be used. A glance at column (3) of Exhibit 4.8 shows that a large majority of the countries covered by the case-studies use census EAs as the basic frame units or building blocks.

Census EAs have several advantages as basic frame units: they usually cover the entire country, they are defined within lower-level administrative subdivisions, measures of size are available, and they tend to be fairly uniform in size. Maps are often available, including base maps that show the location of EAs within each administrative subdivision and maps or sketches of individual EAs, often with detail added by census enumerators. One cannot count on all EA sketches being of adequate quality for survey use or even on all of them being available, but deficiencies can be identified and remedied for EAs included in a master sample or one-time sample. EAs may be too large to serve as the final area units for a particular survey design, but the primary purpose of the MSF is to provide the frame for a sample of PSUs; additional stages of sampling can be used in the sample designs to produce area units of optimum size. All things considered, census EAs, when available, are usually the best choice as basic frame units.

Units smaller than census EAs have some advantages and it may be desirable to use them as basic frame units provided the needed inputs are available and can be incorporated in the MSF without significantly increasing the costs of its development and use. The case-studies provide two examples:

- o In Botswana, the primary stratification for household and agricultural surveys is ecological, i.e., based on land use. There are five strata: towns, villages, lands, freehold farms and cattle posts. Census EAs in towns did not cross town boundaries, but elsewhere EAs sometimes contained a mixture of the four remaining land use categories. Therefore, census EAs were used as frame units for the town stratum, but EAs not in towns were subdivided into "blocks", each limited to one of the four land use categories. This subdivision was initially made for the purpose of current agricultural surveys and was subsequently exploited in developing the household survey MSF. PSUs in the four strata outside of the towns were to consist of blocks or groups of adjacent blocks.
- o In Thailand, blocks within census EAs were used as basic frame units for the municipal area (urban) component of the MSF. Reasonably good street maps are available for all municipal areas. Prior to the population census, EAs and blocks within EAs were identified and numbered on the census maps. On the census EA listing form, a block number was recorded for each housing unit, and counts of households and population by block were obtained manually following the census enumeration. Survey PSUs in the municipal areas were to consist of blocks or groups of adjacent blocks.

As mentioned earlier, inclusion of more than one kind of unit in an MSF with a hierarchical structure leads to greater flexibility in developing an optimum design for each survey. It also creates the possibility of using designs for which both the PSUs and SSUs can be selected from the MSF. Inclusion of units consisting of groups of EAs (or other basic frame units) can be either explicit or implicit, i.e., the frame may or may not include separate records for each of the larger units. In the latter case, the frame should be designed so that the larger units can be formed easily whenever it is decided to use them as PSUs in a survey.

The possible <u>content</u> of the records for the frame units has been discussed earlier in this chapter (see subsections A,5 and C,1). For convenience, eight categories of information that may be included in frame unit records, shown in Exhibit 4.1, are listed below:

IDENTIFIERS	ο	Type of unit code
	ο	Primary identifier
	ο	Secondary identifiers
	ο	Links to higher level units
UNIT CHARACTERISTICS	ο	Stratification variables
	ο	Measures of size
OPERATIONAL DATA	0	Sample usage indicators Undate codes
	U	opdate codes

All of these items are potentially valuable and should be included in the frame record to the extent that data of acceptable quality are available and can be included at a reasonable cost. Planning for frame development as part of the overall census processing operations will help to keep the costs low.

The primary identifier is an indispensible item in the frame unit record. The following guidelines should be considered in developing a system of primary identifiers:

- o Fully numeric identifiers should be preferred to names or alphanumeric codes.
- o The identifier should uniquely identify the basic frame unit and all higher-level units in the hierarchical structure. A fixed number of digits should be assigned to each type of unit and that number should be large enough to meet all requirements. For example, if each province has 9 or fewer districts, one digit will be sufficient for the district portion of the code; but if one province has 10 or more districts, two digits will be needed (see also the following guideline).
- o The numbering system should allow for updating when changes occur in the units at any level. If the order of numbering at any level is alphabetical or geographical, it may be desirable to leave gaps to allow for the insertion of new units created by splits or other changes.
- o Consider the possibility of adding check digits (error-detection codes) to the basic frame unit identifiers.
- o If the preceding requirements are satisfied, rely on existing geographic coding systems.

The <u>choice of medium</u> for the MSF -- hard copy or computerized -will depend on the statistical office's overall capabilities for data processing. A computerized MSF is to be preferred if the computer facilities and personnel are judged capable of developing and using it, especially if the MSF is expected to be used to select independent samples for several different surveys with varying design requirements.

In practice, sample selection from most of the MSFs described in the case-studies has been done manually, the major exception being Australia, where the sample selection operations, at least at the initial stages, are largely computerized. The MSF used by Sri Lanka was a computer listing of identifiers and data for census EAs from that country's 1981 census of population; however, sample selection and subsequent updating of the MSF was done manually. In spite of the current situation, there is much to be gained by aiming at greater computerization of MSF development, use and maintenance. Continuing improvements in hardware and software should make this a realistic possibility.

With respect to <u>auxiliary materials</u>, the key question is what maps should be associated with the MSF. The guiding principle is to rely primarily on materials already available from the most recent census or other sources. The creation of new maps or updating and enhancement of existing maps is expensive, requiring special skills, and should be limited as much as possible to areas where the new materials are needed for sample survey purposes.

Census base maps, showing the location of census EAs in relation to administrative subdivisions, would be a desirable adjunct to an MSF in which the basic frame units are census EAs. These maps will be needed to help the survey field staff locate the sample EAs assigned to them. If maps or sketches of individual EAs were prepared for or during the census, these should certainly be preserved for subsequent use in surveys. Preservation of sketch maps may be facilitated by making them an integral part of the census listing form.

Whatever kinds of maps are made an adjunct to the MSF, attention should be given to the need for linkages between the frame unit records and the maps. There should be sufficient correspondence in the primary and/or secondary identifiers for the frame units to be easily located on the maps.

Other useful auxiliary materials, as already mentioned in subsection A,6 of this chapter, are summary tabulations of the number and characteristics of various kinds of frame units and, especially for computerized frames, a record layout for each type of frame unit and documentation of the sources and definitions of all record fields. These elements are indispensible for effective use of the frame.

4. Preparation of a schedule for initial development of the MSF

As for any complex operation, it is important to have a detailed schedule for the development of an MSF. The schedule should list specific steps in planning, operation and evaluation, and show the proposed starting and completion dates for each step. With such a schedule, the project manager can monitor progress and make adjustments when necessary.

The initial project schedule should include the preparatory steps already discussed in subsections 1 to 3 of this section. It seems more appropriate, however, to discuss the full schedule at this point, because the nature of the operational and evaluative steps depends to a considerable extent on the decisions on key characteristics of the MSF, which have just been discussed in subsection 3.

The presentation here assumes the ideal situation, i.e., that the MSF inputs will come mainly from a population census and that planning for the MSF will have started prior to that census. In other situations, obviously, the nature and sequence of activities will have to be adapted to fit the particular circumstances.

Exhibit 4.9 lists the steps in the initial development of an MSF, given the ideal situation just described. The steps are listed in their approximate time sequence; however, not every step must be completed before the next one begins. For example, the system design for the MSF (item B.3.a) might begin even before the census enumeration is over. However, a system design, at least in preliminary form, must be available before undertaking a small-area test (item B,3.b).

Steps A,1 to A,3 in Exhibit 4.9 have already been discussed in detail in the three preceding subsections. The purpose of step A,4, inputs to census procedures, is to incorporate specific features into the census forms and procedures that will lead to the census outputs needed for the construction of the MSF (selected census outputs are, of course, MSF inputs).

Several relevant aspects of the census enumeration procedures have already been discussed. Features that might benefit the MSF and the household surveys based on it include:

- Division of census EAs into smaller areas for which separate counts of population or housing units would be obtained, e.g., the blocks used in municipal areas in Thailand.
- o Design of the EA listing form and procedures to facilitate quick counts of population and housing units or households.
- Development of an EA numbering system compatible with future MSF requirements.

Exhibit 4.9 Steps in the initial development of an MSF

- A. Pre-census phase
 - 1. Determine general objectives
 - 2. Broad inventory and evaluation of potential inputs
 - 3. Decisions on key features
 - 4. Inputs to census procedures
 - a. Enumeration
 - b. Processing
- B. Post-census phase
 - 1. Assemble inputs
 - a. Lists and data
 - b. Maps
 - 2. Evaluate census inputs
 - 3. Establish frame
 - a. System design
 - b. System test
 - c. Assemble full frame
 - 4. Document frame

.

- a. Tabulate data for frame units
- b. Collect or prepare record layouts and specifications
- 5. Begin use for sample selection

- o Development of a master reference file of EAs, preferably computerized, that can be used to ensure that census inputs to the MSF are complete.
- Assignments, to census enumerators, of specific responsibilities for enhancement of EA maps or preparation of separate sketches, with features that would facilitate future use of these materials for subsampling EAs.
- o Establishment of controls to ensure that EA maps or sketches (or copies thereof) will be available as an adjunct to the MSF.

To help decide which features should be recommended for inclusion in the census processing, an important decision is needed. Should the counts for EAs and other areas to be used as frame units be produced manually or by computer? Typically, population census results are produced in three stages, as follows:

<u>Stage 1</u>. Hand counts of population and households by EA, often made in local offices, and aggregated to produce totals for administrative divisions and subdivisions and the entire country.

<u>Stage 2</u>. Computer tabulations of a limited number of census items, with separate totals by EA, which serve as building blocks for compiling totals for various kinds of administrative and statistical areas.

<u>Stage 3</u>. Computer tabulations of the full range of census items, but with less geographic detail. Frequently these tabulations are based on a sample of census returns.

In this context, the question is: Should the lists of frame units and the corresponding data for the MSF be based on outputs from Stage 1 or Stage 2 of the census processing?

Stage 1 outputs are likely to be available anywhere from several months to a year or more before the Stage 2 outputs. This may be an important consideration if the goal is to produce an MSF for the first time, rather than to update or replace an existing one. In some respects, the Stage 2 data may be of better quality, but with adequate checks and controls, the Stage 1 counts should be sufficiently accurate to use as measures of size. The costs of using Stage 1 data are likely to be greater, especially if the MSF is to be computerized. If Stage 2 data are used, the needed MSF inputs, in the form of a printout or a computer-readable file, can be produced at a relatively low cost.
Once the decision on whether to use Stage 1 or Stage 2 census outputs is made, the schedule for the post-census phase of the MSF development can be laid out in detail, with tentative dates. The main steps are shown in Part B of Exhibit 4.9. Most of these steps have been discussed earlier and do not need elaboration here. Methods of evaluating proposed inputs (step B,2) were discussed in subsection 2 of this section. The development of a system design for the MSF (step B,3,a) should include detailed plans for record formats (hard copy or computer), file structure and organization, and a system of primary identifiers. Prior to full-scale assembly of the MSF (step B,3,c), it is strongly recommended that the system design be tested and evaluated (step B,3,b) in at least one area, say a state or province. The test would consist of assembling the MSF for the selected area and using it to test procedures that will be used on the full MSF to tabulate data for frame units (step B,4,a) and to select samples (step B,5). The results of this test would be used to make improvements both in the design of the MSF and in the computer programs or clerical procedures developed for using it.

5. Development of a plan for updating the MSF

As time passes, changes in the definitions and characteristics of the units included in an MSF are certain to occur. To what extent should these changes be reflected in the MSF?

As in the preceding subsections, it is assumed that the country is following the "ideal" pattern for MSF development and maintenance, i.e., a full update of the MSF after each population census and adjustments as needed during the intercensal period (see Exhibit 4.6).

A "full update" would usually mean replacement of the existing MSF by one based entirely on the current census outputs. When this is done, the selection of samples from the new MSF normally proceeds without reference to prior samples. Special techniques may be necessary if it is desired to retain, during the transition to the new MSF, a survey design with partial rotation of the sample between rounds, or to maximize overlap between new PSUs and PSUs in a master sample based on the old MSF. However, these issues do not generally come up in the household survey programmes of developing countries and they will not be discussed further in this study. <u>This subsection will cover interim</u> adjustments to an existing MSF during its period of use.

Changes in frame units are of two kinds:

(1) Changes in frame unit boundaries. These changes primarily affect administrative subdivisions. New subdivisions are created and the boundaries of existing ones are changed.

(2) Changes in frame unit characteristics. Such changes include:

- o For units that are administrative subdivisions, name changes.
- For any kind of frame unit, changes in the identification of higher-level administrative subdivisions in which the unit is located.
- o Changes in characteristics used as stratification variables, e.g. urban-rural classification.
- o Changes in characteristics used as measures of size, e.g., population and number of housing units or households.

To what extent should the MSF be revised to reflect these kinds of changes? Answers to this question depend on how the MSF is being used and on how those uses are likely to be affected by the changes. For example:

- o If there are changes in the definition of administrative subdivisions or statistical areas for which separate survey estimates are to be made, these changes clearly should be reflected in the MSF.
- o If the name of an administrative subdivision is changed, the new name should be put into the MSF. The change would not be essential for sample selection purposes, but field workers will need to know the new name of the unit if it is selected.
- o Changes in stratification variables and measures of size do not necessarily have to be reflected in the MSF, but adjustments are sometimes desirable if they can be made at a reasonable cost. Such changes do not result in the selection of biased samples but do result, if no adjustments are made, in loss of efficiency, i.e., larger sampling errors for a fixed sample size.

Any change affecting the boundaries of frame units that are used as PSUs or to form PSUs must be reflected in the MSF in some way. To take a simple case, suppose the frame units are villages and a village has been split to form two villages. One possibility would be to delete the record for the old frame unit and substitute two new ones. Another possibility would be to retain the record for the old unit, but to delete the old village name and add the two new names to the record.

The appropriate strategies for updating an MSF depend, in large part, on the design of the IHSP for which it is being used. If a new, independent sample of PSUs is to be selected for each survey and survey round, then all information on changes that is readily available should be incorporated in the MSF and there would be no particular need for maintaining links between old and new units. On the other hand, if a particular sample of PSUs is to be used, perhaps with partial rotation, over a long period, different strategies are needed for adjusting the MSF. Frame units that have been affected by changes in definition or measures of size must be carefully identified or "tagged" so that supplementary samples of new PSUs and those with unusually large growth can be selected and combined with the existing sample in an unbiased manner. It will be convenient to refer to the first of these two design strategies as the <u>sample replacement</u> strategy and the second as the <u>sample revision</u> strategy.

These general ideas can be made more concrete by taking some examples from the case-studies. Surprisingly, for the majority of the countries for which case-studies were prepared, the available documentation made no mention of procedures or plans for updating the MSF, even though in most cases, surveys were to be conducted over periods of several years. This observation does not prove that insufficient attention is being given to the problem of changes in frame units, but suggests that it is perhaps the case.

Three countries -- India, Sri Lanka, and Thailand - follow the sample replacement strategy, i.e., they periodically update the MSF and select new, independent samples of PSUs. Only one of the countries included in the case-studies -- Australia -- follows the sample revision strategy. Brief descriptions of the procedures used in each country follow:

<u>India</u>. A new, independent sample of rural census villages and urban census blocks is selected for each annual round of the National Sample Survey. According to case-study reference 2, "As there are frequent changes in the boundaries of census blocks, the NSS updates them in a phased manner during the intercensal period."

Sri Lanka. Following the first of a planned series of annual household surveys, the census-based MSF was updated to reflect changes in the definition and characteristics of the frame units. Frame changes included deletion of some frame units, creation of some new units, and changes in measures of size for units with substantial amounts of new housing. In addition, identifiers of higher-level administrative units were changed to reflect the creation of a new district. This change was essential, because the surveys are designed to produce separate estimates for each district. A new, independent sample of PSUs was selected from the updated MSF, to be used in the second and third annual household surveys. Following the third survey, it is likely that the MSF will be fully updated, using the result of a scheduled mid-decade population census. Thailand. The frame units for the rural (non-municipal) part of the country are villages. The rural component of the MSF is a list of villages which is periodically updated on the basis of information obtained from the Department of Local Administration, Ministry of Interior. The number of villages has increased at a rate of about two percent annually in recent years. The village frame is used for household surveys, decennial censuses of population, and annual surveys in which information on village characteristics is obtained from all villages. Population and household counts are available from the censuses and the annual village surveys. For the household surveys, a new independent sample of villages is selected each year, using the MSF as it exists at the time of selection.

Australia conducts a population census every five Australia. years. Following each census, a new MSF is constructed and a master sample selected for use during the subsequent five-year period. The master sample PSUs, which are census EAs or groups of EAs, are used for a Monthly Population Survey and for various supplementary The MSF measures of size are updated surveys during this period. every six months, primarily on the basis of building permit information collected for use in a construction statistics program. When some of the PSUs in a stratum have grown by unusually large amounts, new sample PSUs are added to the master sample, using a sampling procedure that reflects their growth in an unbiased fashion. A separate component of the MSF is maintained for special dwelling places. The list of special dwelling places is updated at least every six months, using information supplied by state offices of the Australian Bureau of Statistics. The updates include new special dwelling places, deletions and changes in estimated occupancy for units already listed. Sampling procedures for the special dwelling unit stratum reflect these changes.

A plan for updating the MSF should be developed as early as possible during the MSF development process, because it may have important implications for the structure of the initial version. The design of record formats and the system of numeric identifiers should allow for changes that will be required when the MSF is updated.

The principal elements to be included in the update plan are:

o Sources of change information to be used for updates. Possible sources include: other censuses (e.g., a census of agriculture might provide current housing unit counts for rural EAs), administrative records, and field-work carried out specifically for use in updating the MSF. One caution is necessary: field-work conducted in current sample surveys should not be used as a source of MSF change information because this could lead to differential treatment for sample and non-sample PSUs, resulting in future selection biases.

- o Criteria for the creation of new frame-unit records.
- Criteria for deletion of records. A subsidiary question is whether deleted records should be preserved in a separate file or deleted altogether.
- o Criteria for changes to existing records.
- o The frequency with which various types of updates are to be made. One option is to incorporate new information whenever it becomes available. Another is to incorporate all change information at specified intervals, e.g., annually. Frequency might vary for different kinds of changes.
- o Strategy for recording changes over time. One possibility is to maintain only the most current version of the MSF, reflecting all additions, deletions and changes to date. A second option is to design frame-unit records which can show prior as well as current information, at least for selected items. A third option is to keep a tape copy or hard copy of the MSF as it existed at the time of each use for sample selection. The third option assures the availability of universe and sample information that may be needed for estimation based on data for a particular sample.

Decisions about each of these elements will depend to a considerable degree on whether the country adopts the sample replacement or the sample revision strategy to maintain design efficiency in the face of changes in the target population. The relative merits of these strategies will be discussed in Chapter V.

The plan for updating should be designed to make only those changes in the MSF that are essential or desirable because of the nature of the uses that are being or will be made of the MSF. The benefits from desirable changes must be weighed against their costs.

6. Adaptations for less than ideal circumstances

The five preceding subsections have described the development and maintenance of an MSF in the ideal situation where planning begins prior to a population census and actual construction of the initial frame starts when the census outputs are available. This subsection will consider some alternatives when development begins at a different stage of the census cycle or when no census is planned or has been taken. General comments will be followed by some illustrations from the case-studies.

Several kinds of problems can be encountered in these less than ideal circumstances. If no census has been taken or is planned, the inventory and evaluation of potential frame inputs from sources outside of the statistical system takes on added importance. It is likely, although not certain, that an MSF constructed in these circumstances will lack some of the desirable qualities discussed earlier in this chapter.

If there has been a recent census, but the outputs needed for an MSF were not considered sufficiently in planning and conducting the census, other problems may exist. Lists of EAs and population and household or housing unit counts for EAs may not be easily accessible. Some cartographic materials, such as base maps and enumerator sketches, may be missing or of poor quality. If some time (say 3 or 4 years) has elapsed since the census data collection, various kinds of changes will have occurred in the definition and characteristics of potential frame units. In this last case, the problems to be dealt with are similar to those faced in connection with updating an MSF, as discussed in the preceding subsection.

Whatever the deficiencies in the materials available for an MSF, they must be taken into account in developing the design for an IHSP. If no suitable frame materials are available for some areas or population groups, these areas or groups may have to be omitted from survey populations until adequate frame materials can be developed to cover them.

It may be necessary to make a choice between small frame units which are not well defined and larger units whose boundaries are clearly defined and mapped. While the smaller PSUs might be closer to the optimum size when costs and sampling errors only are considered, the larger PSUs might be preferred in order to reduce the likelihood of significant coverage errors. With the larger PSUs, additional stages of sampling can be introduced and good quality subsampling frames developed only for the sample PSUs (see Section E of this chapter for further discussion of secondary sampling frames).

Lack of adequate frame materials may also necessitate cutting back on the data requirements for the IHSP, especially in the designation of geographic areas for which separate survey estimates are to be developed. The need to use large PSUs might make it impractical to use a survey design that would produce reliable estimates for each state or province: national or at most regional estimates may be all that is feasible at the start.

The case-studies provide several examples of adaptations to limited availability of inputs for the development of an MSF:

Ethiopia. Prior to the period covered by the case-study, no census had been conducted in Ethiopia. The smallest administrative subdivisions of the country are Urban Dwellers Association and Farmers' Association areas. It is estimated that there are more than 1,300 of the former and 25,000 of the latter. The average size of a Farmers' Association area is about 250 households. During a two-month period in 1980, a list of 18,989 Farmers' Associations was compiled. The list covers 12 of the country's 14 regions and 419 sub-provinces in 77 of the 85 provinces within the 12 regions. This list, which served as the MSF for Ethiopia's Rural Integrated Household Survey Programme, included an estimate of the number of members in each Farmers' Association. The documentation on which the case-study was based does not say whether maps showing the location and boundaries of the Farmers' Association areas were available, nor does it discuss the stability of these units over time.

<u>Morocco</u>. The basic frame units for urban areas were census EAs and the PSUs consisted of groups of EAs. In rural areas, <u>douars</u> had been used as the unit for census enumeration. <u>Douars</u> are social units with a head or chief and do not always have fixed boundaries; therefore they were not considered suitable for use as frame units. Consequently, communes, the smallest administrative subdivisions with clearly defined boundaries, were chosen as rural frame units. Most rural PSUs for the master sample were individual communes; in a few cases smaller communes were combined to form PSUs. Rural PSUs selected for the master sample were sent to the field for mapping and subsequently divided into segments, with an average of 1,000 households, for use in the next stage of sampling.

<u>Nigeria</u>. Census EAs were the basic frame units for the MSF that was developed for Nigeria's National Integrated Sample of Households (NISH). The census was conducted in 1973, but the MSF for NISH was developed several years later, for use in surveys starting in 1981. At that time, no counts of population, housing units or households were available. The EAs were defined on sketch maps; however, sketch maps were unavailable for an unspecified proportion of the EAs.

A master sample of EAs was selected in each of Nigeria's 19 states for use in NISH surveys. Presumably because no measures of size were available for EAs, the sample was selected in two phases. In the first phase, a stratified sample of 200 EAs was selected in each state, using equal probabilities of selection within strata. Household listings were then prepared for each of the sample EAs, following which a subsample of EAs was selected in each stratum with probability proportionate to number of households (farm households in the rural strata).

This is an interesting way of trying to compensate for the absence of measures of size; however, it is not clear whether the double sampling procedure was more efficient than the alternative of choosing a somewhat larger sample of EAs with equal probability in a one-phase selection process. Evaluation of these alternatives would require more information on variances and costs than is provided by the documentation.

The MSF developed for Saudi Arabia's Multipurpose Saudi Arabia. Household Survey was a list of emirates, along with their population counts from the country's 1974 census of population. The emirates are the smallest administrative units with definable boundaries. Enumerator assignments for the 1974 census were established within emirates. but apparently were not considered sufficiently well-defined for use as frame units. Emirates were classified as metropolitan, urban and rural. Rural emirates with fewer than 5,000 settled population were combined with adjacent emirates to form The MSF contained a total of 137 PSUs: 10 metropolitan, 32 PSUs. urban and 95 rural. Sample PSUs were subdivided into segments groups of municipal blocks in metropolitan and urban areas and groups of villages in rural areas -- for the next stage of The approach used in Saudi Arabia was quite similar to sampling. that used for the rural sector in Morocco; establishment of relatively large PSUs and development of suitable secondary sampling frames limited to the sample PSUs.

Thailand. The municipal area (urban) component of the MSF for Thailand. The municipal area (urban) component of the MSF for Thailand's household surveys is based on the most recent census of population. The frame units are blocks, which are defined areas within census EAs. Census counts of population and households are available for the blocks. The rural component of the MSF is a list of villages, which is updated annually on the basis of information from the Department of Local Administration, Ministry of Interior. New villages are created, mostly by splitting existing villages, at the rate of about 2 percent annually. Villages do not have officially-defined boundaries. Population and household counts are available from the 1980 census (for villages that existed at that time) and, for most villages, from the latest annual village survey

In its early household surveys, Thailand used larger administrative subdivisions, called amphoes, as rural PSUs and villages as SSUs. Starting early in the 1980s, villages were adopted as rural PSUs for the continuing labour force survey. A new sample of villages was selected annually, using equal probability selection within strata. It was proposed in 1983 that villages be selected with probability proportionate to size in subsequent surveys, using measures of size developed primarily from the annual village surveys. In the longer run, it can be hoped that the development of more detailed maps covering rural areas will permit the use of PSUs and SSUs with well-defined boundaries, thus making it more feasible to consider using sample designs that retain selected rural PSUs for more than a year.

Although the population census has been identified in this chapter as the ideal source of inputs for an MSF, one of the case-studies

e

describes a high-quality, extensively-used master sample selected from an MSF that did not rely on census inputs, namely, the United States master sample of agriculture. The primary input used to construct the frame for the master sample of agriculture was a set of detailed up-to-date county highway maps which were not only rich in features that could be used o delineate area segments but also contained symbols showing the residences of farm operators. With these maps it was possible to develop and make a list of well-defined, relatively stable area frame units, with appropriate measures of size, i.e., measures based on the number of farms in each area unit. The master sample of agriculture was a sample of these area units, which were known as "count units". The sample count units were subdivided into smaller area segments, again making use of the county highway maps for this purpose.

E. Secondary sampling frames (SSFs)

In a multi-stage survey design, every stage of sampling requires a frame. The MSF can be used for the first stage of sampling and sometimes for the second stage, but for additional stages other frames must be developed. These frames, which are needed only for the sample PSUs (or SSUs if the MSF is used for the first two stages of selection), will be referred to as secondary sampling frames (SSFs).

Like any frame, an SSF can consist either of area or list frame units. It is also possible, at a particular stage of sampling, to use a frame consisting of both kinds of units. Exhibit 4.10 shows, for some commonly used multi-stage designs, the kinds of frame units used for each stage of sampling. Villages, housing units (HUs) and households (HHs) are assumed to be list units; all others shown are area units. Normally, list units are used only at the final stage of sampling, the exception being the use of villages as PSUs in design 5. At stage 2 in design 5, the frame units could be area segments in villages for which suitable maps were available for use in forming segments and housing units in other villages.

Subsections 1 and 2 of this section discuss the development, updating and properties of area and list SSFs respectively. Subsection 3 discusses current practices, as shown by the case-studies, and presents some general recommendations.

1. SSFs with area units

The construction of an SSF with area units requires, for each sample PSU or SSU. a map with sufficient detail to divide it into smaller areas, according to specified criteria. These small areas will be referred to as <u>area segments</u> or simply segments (other terms, such as blocks, zones or chunks are sometimes used). The process of subdividing PSUs or SSUs into area segments will be referred to as <u>segmentation</u>.

Design number	Stage 1	Stage 2	Stage 3	Stage 4
1	Census EA	ни/нн		
2	Census EA	Area segment	ни/нн	
3	Census EA	Area segment		
4	Administrative subdivision	EA	Any of the co used in stage of designs 1	ombinations es 2 and 3 to 3
5	Village	HU/HH or area segment		

Exhibit 4.10 Frame units for some typical multi-stage designs

Notes:

1. For all designs, stage 1 selection is assumed to be from an MSF. For design 4, stage 2 selection may also be from an MSF.

2. When the units for the final stage of sampling are area segments, an HU or HH frame will be prepared but not sampled.

3. EA - enumeration area, HU - housing unit, HH - household.

.

The purpose of segmentation is to reduce the amount of listing required. Suppose the target sample size within each sample SSU were 20 housing units and that the average SSU contained 200 housing units. Without segmentation, it would be necessary to list 200 housing units per segment. With segmentation, this number could be reduced substantially. For example, if segments of average size 50 could be formed, the listing requirement would be reduced by 75 percent.

The criteria for the segments to be formed should cover three aspects: definition, size and number. With respect to definition, the main objective is to use, as much as possible, stable physical features, such as streets, roads, railroads, rivers and streams, for segment boundaries. If some of the boundaries of the PSU or SSU are imaginary lines, e.g., boundaries of administrative subdivisions, they will of course have to be used for segments that share those boundaries.

The <u>size</u> of area segments is usually measured by the actual or estimated number of households or housing units they contain. A minimum size is almost always established: there is nothing to be gained by further subdivision below this level even if suitable features are available as boundaries. Usually an upper limit is established as well, so that all segments are expected to be within a specified range. The range may be broad or narrow, depending on the sample design. If all of the area segments are to be used as "take all" segments (segments in which all housing units or households are included in the sample), then a narrow size range is desirable, but if some or all of the selected segments are to be listed and samples of housing units or households selected, a wider size range can be tolerated.

For designs that provide for self-weighting samples within strata, the specifications for segmentation may also require that each sample PSU or SSU be divided into a designated number of area segments or, alternatively, into area segments whose measures of size, in terms of the number of ultimate clusters, add up to the PSU or SSU measure of size. To illustrate the second alternative, consider a design for which the target ultimate cluster size is 10, i.e., it is planned that, on the average, one cluster of 10 housing units will be selected in each PSU. Suppose a PSU has an estimated 82 housing units and has been assigned measure of size 8. The segmentation operation has produced area segments as follows:

Segment	Estimated	Preliminary	Final
number	HUs	measure	measure
1	10	1	1
2	14	1	1
3	24	2	3
4	20	2	2
5	14	1	1
Total	82	7	8

The preliminary measures of size did not add up to 8, so an adjustment was made in segment number 3, where it would have the smallest effect on the ultimate cluster size. To maintain the self-weighting sample, segments 1,2, and 5, if selected, would be treated as take-all segments. Segments 3 and 4 would require listing and further sampling, at the rates of 1 in 3 and 1 in 2, respectively.

A critical factor in deciding whether it is feasible to develop and use SSFs with area units is, of course, the nature and quality of the maps available for segmentation. Consider an imaginary country which has good census sketch maps for nearly all urban census EAs so that there will be little difficulty in dividing all urban EAs into segments of size 20 ± 5 HUs. For the rural census EAs however, the nature of the sketch maps is such that further subdivision would be difficult. The rural EAs are mostly in the size range 50 to 200 HUs. A possible design would be as follows:

Urban EAs. Using the census sketch maps, divide the sample EAs into take-all segments of average size 20.

Rural EAs. List HUs in each sample EA and sample to get the desired number or proportion of HUs.

An alternative in this example would be to do further field work for all sample census EAs in the rural stratum with more than, say, 100 HUs to develop sketch maps that could be used to divide them into two or more area segments. Two stages of sampling could be used in each of these EAs: selection of an area segment followed by HU listing and sampling in the selected segment.

As illustrated by this last alternative, some field-work to produce maps or sketches suitable for segmentation of PSUs or SSUs is a possibility. The costs of doing the field-work must be balanced against the expected reduction in listing costs (including updating).

Office and field staff who do mapping and segmentation should have considerable training and experience. A detailed description of mapping techniques is beyond the scope of this technical study. Two useful references are Poplab Manual No. 1, <u>Mapping and House Numbering</u> (Cooke, 1971) and the U.S. Census Bureau Training Document, <u>Mapping for Censuses and Surveys</u> (U.S. Bureau of the Census, 1978).

If desired, SSFs with area units can be used with little or no updating, insofar as the definitions of the units are concerned, for several surveys or survey rounds over an extended period. Physical features used as boundaries seldom disappear. Changes are, of course, possible in administrative subdivision boundaries that are also being used as segment boundaries. When changes of any kind make it necessary to restructure segment boundaries, it is important to use a procedure that is not influenced by knowledge of which frame units have actually been selected. Changes should not be based solely on information obtained as the result of field-work in sample units. Preferably the person who redefines the frame units should have no knowledge of which ones are currently included in surveys.

SSFs with area units have the important advantage that they require less updating than list frames, especially when the segment boundaries are predominantly physical features. The initial cost of creating area SSFs in a specified set of PSUs is likely to be less than the cost of doing a complete listing of housing units or households in the same set of PSUs. If the SSFs are to be used more than once, the cost advantage of the SSFs with area units becomes even greater, because they require very little updating, whereas list frames must be updated periodically.

Creation of area SSFs, however, requires skills that may not be as readily available as those required for listing operations. An effective quality control system must be established to assure that the segmentation procedures are carried out according to specifications. Much depends on the target size or size range established for the area segments. In many areas the pattern of habitation is such that it is simply not feasible, even for highly-trained field workers, to delineate clearly-defined segments with as few as 5 or 10 housing units. This feasibility factor places limits on the use of area SSFs for successive stages of sampling in multi-stage samples. Just where this limit occurs must be determined empirically for different countries and different types of areas within a country.

2. SSFs with list units.

This subsection will be concerned primarily with SSFs whose frame units are groups of persons -- usually called households -- or structures -- variously referred to as housing units, dwelling units or living quarters. Before discussing SSFs with these kinds of units, however, brief mention should be made of some other kinds.

<u>Villages</u> which do not have defined boundaries should be considered list units rather than area units. If a complete list of villages for the entire country is used for sampling, it is considered to be an MSF. An alternative design, however, is to use higher-level administrative subdivisions as PSUs and establish SSFs consisting of village lists only for the sample PSUs. Other types of list frames that might be established only in sample PSUs are lists of institutions and other special dwelling places. Sampling and operational considerations frequently make it desirable to sample the population in these units separately from those living in regular dwelling places.

In some instances it may also be desirable to sample from SSFs that are lists of <u>persons</u>. If a sample of institutions or special dwelling places has been selected, application of the principles of optimum sample design usually leads to sampling of individuals, at least in the larger units. If such a list already exists, e.g., a roster of resident officials and inmates in a penal institution, it may be used directly for subsampling if it appears to be complete. Otherwise, field-workers will need to prepare the lists.

Returning now to the main topic of this subsection, SSFs consisting of lists of households, housing units or living quarters, a key question is: What kinds of residential units should be preferred as list frame units? This question was addressed in detail in the <u>Handbook of</u> <u>Household Surveys</u> (United Nations, 1984) and the main points are worth repeating here:

In a household survey, the natural ultimate sampling unit might be the household. Using the definition recommended by the United Nations for population censuses, a household would comprise either an individual who makes provision for his or her own food or other essentials for living or a group of two or more persons living together who make common provision for food or other essentials for living.

One problem with using households as the ultimate sample units is that they lack permanence and may change between the time of sample selection and the start of data collection as a result of the mobility of some or all of the members. Moreover, households are not readily identifiable from external features but will usually require inquiries to establish their identity. A more permanent type of unit, which can usually or often be identified by external observation, is the housing unit or, more broadly, living quarters. According to the United Nations housing census recommendation, living quarters are separate and independent places of abode intended for habitation or not intended for habitation but occupied as living quarters at the time of the census. Where living quarters is specified as the ultimate sampling unit, all of the households living in a selected unit - and there could be more than one - are covered by the sample.

In some countries, where extended families living in compounds are common, neither of the above concepts (household or living quarters) may be feasible. It may often be necessary, in these situations, to consider the entire compound as the ultimate sampling unit.

The position taken by the World Fertility Survey (1975) in its <u>Manual on Sample Design</u> is similar, although not quite as clear-cut. The surveys taken under the WFS programme were planned as one-time, <u>ad</u> <u>hoc</u> surveys, with households and individuals as the elementary units. In discussing alternatives for list sampling frames, the following points are made (WFS, 1975, pp. 24-26):

- o Existing household lists should rarely, if ever, be used.
- o (Direct quote) "In practice, a fertility survey will nearly always have to depend on a household listing operation carried out specifically for the survey, unless a dwelling sampling frame can be used as a substitute."
- Under many conditions, existing or specifically-constructed dwelling lists are acceptable substitutes for household lists and can be obtained or developed with less effort.

The excerpts from these two manuals make it fairly clear that households should only be used as list frame units in certain rather Specifically, household list frames might be limited circumstances. preferred when two conditions are met: first, that all of the surveys which will make use of the same frames are to be conducted during roughly the same relatively short period and, second, that there will be only a short interval between the time the households are listed and the time when sample households selected from these listings will be interviewed. If these two conditions apply, the household as a listing unit offers some advantages in terms of sampling efficiency. At a given time, households tend to be less variable in size than housing units. Also, when households are used as frame units, auxiliary information, such as number of persons in household or household income class, can be collected at the time of listing and used to select samples that are somewhat more efficient for the purposes of the survey.

However, whenever the conditions of the preceding paragraph do not apply, strong arguments favour the choice of housing units over households as the frame units for SSFs. In particular, if the same SSFs are to be used in more than one survey round or for separate surveys conducted at different times, the greater stability of housing units argues strongly for their use.

This preference for housing unit SSFs should not be taken to mean that they can be used indefinitely without updating. However, updating of housing unit lists can be a relatively simple and inexpensive process, provided the housing units included on the initial list have been adequately identified on the listing forms and maps. Updating can be further facilitated, if local conditions permit, by marking or labelling each housing unit with its survey identification number. On the other hand, updating of household lists can be a complex undertaking. If many changes have occurred, a complete new listing may be the only practical alternative.

Given this general preference for housing units over households as list-frame units, the remainder of this section will emphasize techniques for constructing, using and updating housing-unit SSFs and will discuss their advantages and disadvantages relative to SSFs with area frame units. Nevertheless, much of what will be said is equally applicable to the construction of household list frames, because it is a fairly common practice, even when households are used as listing units, to use housing units as first-stage listing units and then to list the households associated with each housing unit.

The planning and development of housing-unit SSFs is much like the process of developing an MSF, and it may be helpful to the reader to refer back to Exhibit 4.5, which lists the major steps in that process. Indeed, many of the steps here are very similar, and it should be sufficient to point out features that have special importance in connection with housing-unit SSFs. Exhibit 4.11 identifies the principal steps and key issues in planning for housing-unit listing activities.

As in planning for the development of an MSF, one should start by specifying the purposes for which the housing-unit SSFs will be used. Relevant questions include:

- How many separate samples will be selected from the housing-unit listing for each sample PSU or SSU?
- o How many times and at what intervals will each sample be used?
- o What specific sample selection method will be used to select the housing unit sample(s). The method most commonly used is systematic sampling with random starts, but other methods involving stratification based on housing unit characteristics or clustering of adjacent units are possible.
- o Who will do the listing, who will do the sample selection, and what will be the timing of these operations in relation to the conduct of interviews for the selected housing units? It is theoretically possible to list, sample and interview in a single operation; however, this procedure is usually avoided because of the risk of introducing various kinds of selection bias.
- o Will updating of housing unit listings be necessary? If so, how often and when?
- Would it be desirable, as part of the listing operation, to collect and tabulate a few simple data items for all housing units in the sample PSUs or SSUs?
 At first thought this possibility may seem appealing; however, adding anything but extremely simple data items could substantially increase the time required for the listing operations.

Answers to these questions will guide the subsequent development of the detailed plans and materials needed.

	STEP	KEY ISSUES
1.	Determine general objectives and strategy	o Number and timing of samples to be selected.
		o Sampling pattern(s) to be used.
		o Timing of listing in relation to sample selection and interviewing.
2.	Identify and evaluate available inputs	o Possible use of census listing forms.
		o Evaluation of locally available lists.
3.	Decide on key characteris	tics
	a. Coverage	o Categories of units to be listed.
	b. Frame units	o Clear definition of housing unit needed.
	c. Record content	o Ability to locate units in future is critical.
		o Are screening or stratification vari- ables needed?
	d. Storage and pro- cessing medium	o Generally hard copy only.
	e. Auxiliary materials	o Sketch showing location of housing units.
		o Can identification numbers be attached to units?
4.	Prepare schedule for in-	o Form development.
	Itial Hoting Operation	o Pretesting
		o Training of field staff
		o Quality control
5.	Develop plan for	o Frequency and timing
	арастив	o Method of recording units deleted and added.

Exhibit 4.11 Steps in planning for the preparation and use of housing unit listings

The next step in planning is to identify inputs to the listing operation. In most cases it will be preferable to "start from scratch", i.e., to do a completely new listing for each sample PSU or SSU. However, there are some possible exceptions. If housing-unit listings of reasonable quality are available from a recent census for all or most sample PSUs or SSUs, it may be possible to construct the SSFs by doing field updates of these listings. If census listings are used, their completeness and accuracy should not be taken for granted. Listers should be instructed to look carefully for new or missed units and to check the information recorded in the census for the units that were Dwelling unit listings or lists of structures are sometimes listed. available from local sources. For example, village chiefs may have lists of families living in their villages, and most families may occupy Such lists are even less likely than census separate housing units. listings to satisfy all requirements for the desired SSFs, but, if available, could be used by listers as a starting point in preparing their own listings. However, such locally available lists should never be assumed to be fully complete and accurate.

Like the plan for an MSF, the plan for developing housing-unit SSFs requires decisions on key frame characteristics. The decisions about coverage and frame units are closely related. It is not sufficient simply to specify that all housing units should be listed, or to rely only on the definition of living quarters from the UN Handbook of Household Surveys cited earlier in this subsection. Listers must be given a clear, precise housing-unit definition, reinforced by examples covering borderline cases that are likely to occur (useful guidance is provided in the United Nations Principles and Recommendations for Population and Housing Censuses, 1980a). Should units occupied by inmates and officials of institutions be included in the listings? What about other kinds of collective living quarters, such as hotels, boarding houses, construction camps, etc.? It may be desirable to list these special varieties of living quarters, but to identify them separately from regular housing units.

<u>Record content</u> refers to the information that will be recorded for each housing unit on the listing form. Above all, the listing form must contain enough identifying information for all sample housing units to be readily located and identified by interviewers. The name of the principal occupant by itself is not sufficient, since a different group of persons may be occupying the housing unit when the time for interviewing arrives. Street names and numbers should be recorded, if available, and apartment numbers, if relevant. Other distinguishing features should be included, such as the presence of a shop in the unit (and its name), an unusual tree in front of the unit, the material used for walls or roof, if different from neighboring units, and so forth. Without this kind of information, it will be difficult for supervisors to check the completeness and accuracy of listings and for interviewers to be sure they have correctly identified the sample housing units. Data items for the housing units listed frequently include type of unit (regular or collective), number of households living in unit, number of persons living in unit, whether or not any resident operates one or more agricultural holdings and whether or not any resident operates a non-agricultural enterprise. If the housing unit listings are to be used as SSFs for surveys covering some special groups of population, such as disabled persons, the listings must identify units having one or more such persons. Such items are known as screening items.

The storage and processing medium for housing-unit SSFs is usually hard copy, i.e., the listing forms for the individual PSUs or SSUs. Precautions should be taken against loss of listing forms, since considerable work would be required to recreate them. It may be wise, for example, to keep central office copies of listing forms if one set has to be sent to the field for updating or other purposes.

<u>Auxiliary materials</u> should virtually always include maps or sketches of the areas listed. Sketches may be based on existing maps, or they may be prepared entirely by the listers. In either case, they should show the main physical features of the area and a symbol showing the location of each unit listed, along with its serial number. Further details on the preparation of sketch maps in connection with housing unit listings, with illustrations, may be found in Poplab Manual No. 1, Mapping and House Numbering (Cooke, 1971).

Also in the category of auxiliary materials are serial number labels physically affixed to listed housing units. If local conditions permit, housing units can be labelled at little additional cost. The benefits of labels for checking the listings, interviewing, and updating the listings, both in terms of cost and quality, are obvious.

Finally, as in the case of an MSF, auxiliary materials for housing-unit SSFs should include complete documentation. One part of the documentation will consist of the listing form and the associated instructions and training materials. A second part will be control records and tabulations showing the status of initial listing, sample selection and update operations, and the distribution of the PSUs or SSUs by number of housing units listed.

After the key decisions have been made, the activities needed to prepare for and carry out the initial listing operations should be scheduled. If the listing form or procedures differ in any important ways from those used previously in population censuses or household surveys, they should be tested in a few areas that provide examples of different kinds of housing arrangements. The schedule should include quality control activities, such as office and field checks, by field supervisors, of the listings. A plan for updating the listings should be prepared at the same time. For some IHSP designs, no updating may be needed. Initially, the length of time for which a housing-unit listing is usable without updating is likely to be a matter of judgement; in many areas one year might be a reasonable cutoff. As the survey programme progresses, the period of use prior to updating might be shortened or lengthened on the basis of the observed frequency of changes in housing units. If PSUs or SSUs for which housing-unit listings have been prepared are to remain in sample for more than the designated period, they should be updated at scheduled intervals.

The procedures for updating will depend on how the updated listings are to be used for subsampling. One option is to retain the initial sample and supplement it with a sample (usually selected at the same rate) of new housing units identified in the update. Another option is to select new samples without regard to earlier selections. Various sampling schemes can be used to control the extent to which new samples overlap with earlier ones.

Whatever scheme is to be used to select samples from the updated listings, the listing form and procedures should be designed so that housing units added at each update can be distinguished and deletions, e.g., demolished units, can be readily identified. Depending on the sampling procedure to be used after updates, it may be desirable to provide extra columns on the listing form (assuming that a linear format is used) to renumber units on the updated list.

housing-unit The advantages and disadvantages of SSFs are essentially the opposite of those described earlier for area SSFs. Mapping requirements for listing are less demanding than they are for segmentation. Given a defined sample PSU or SSU, the procedures for preparing a housing unit listing are probably more straightforward and require less training than procedures for subdividing the PSU into area segments. The use of list SSFs permits the selection of more widely dispersed ultimate clusters, and this likely 18 to be more cost-efficient from the sampling point of view unless the cost of travel between housing units within clusters is thereby raised significantly.

On the other hand, the initial and updating costs for housing-unit SSFs may be considerably higher than those for area SSFs. Thus, the longer the period for which a particular sample of PSUs or SSUs is to be used, the greater the advantage of the area SSF over the housing-unit SSF.

3. Discussion and summary

It will be useful at this point to examine the ll case-studies to see how the countries involved have created their SSFs. While it may seem, at first glance, that a bewildering variety of designs is being used, careful analysis turns up a limited number of distinct patterns. The design which makes the greatest use of list SSFs is the one which uses census EAs or equivalent units as PSUs, prepares complete listings of housing units or households for the sample PSUs, and subsamples from these listings. Countries that used this design are: Ethiopia, India, Nigeria, Sri Lanka and Thailand (for the non-municipal area stratum). Since Ethiopia had not had a census, it used farmers' associations in place of census EAs. Thailand used villages in place of census EAs: the villages do not have defined boundaries but are in the same size range as census EAs in most countries.

At the opposite end of the spectrum, the U.S. master sample of agriculture was a two-stage area sample in which the SSUs were area segments having, on the average, about four farms (agricultural holdings). Details on specific uses of the master sample are not available, but in all likelihood most users selected subsamples of these area segments and treated them as take-all segments.

One country, Australia, used census EAs as PSUs or SSUs and divided the EAs into area segments. The area segments in rural areas were designed to be take-all segments, averaging from 6 to 10 dwellings. The urban segments were to be somewhat larger so that the segment listing for each one could be used to select from 4 to 8 non-overlapping systematic samples of dwellings.

Thailand, for its municipal area stratum, used defined areas within census EAs as PSUs. These areas, called blocks, were listed and subsampled. Morocco, for its urban stratum, used area segments as SSUs. The sample PSUs were divided into area segments, averaging 50 households.

Two countries, Morocco (for its rural stratum) and Saudi Arabia, used administrative subdivisions larger than census EAs as PSUs because the census EAs were not considered to be sufficiently well-defined for sampling purposes. In each country, the PSUs were to be divided into smaller area units. In Morocco this was to be done in two stages, ending with segments of about 100 households each. The design for Saudi Arabia called for a single stage of segmentation of sample PSUs to produce area segments with from 100 to 200 households. Sample area segments would be listed and subsampled.

The design described in the Botswana case-study is difficult to classify. Some sample PSUs were treated as take-all segments, averaging about 50 households. For larger PSUs, subsampling was required. In some of them, including all of the urban (town) PSUs, dwellings were selected systematically from the census listings; other PSUs were divided into a specified number of roughly equal-size segments and one of these was chosen at random. The documentation for the case-studies did not follow a standard terminology for describing list-frame units, so it was not always possible to be sure whether their actual definitions were closer to the household or the housing unit/living quarter definitions adopted by the United Nations (cited in subsection 2 of this section). As nearly as can be determined, the 10 countries (excluding the U.S. master sample of agriculture, which did not use list-frame units) divided about equally in deciding whether to use households or housing units. As indicated in the Thailand case-study, a switch from households to housing units was proposed but has not yet been adopted.

Probably the main conclusion to be drawn from these examples is that the choice of secondary sampling frame units in developing countries depends largely on the nature of available maps and the availability of field personnel with the ability to enhance existing maps or to prepare sketch maps for small areas. The discussion of the respective advantages and disadvantages of list and area units for SSFs in subsections 1 and 2 of this section should have made it clear that area units are to be preferred at all stages of sampling, provided conditions are such that well-defined area units can be established at a reasonable The recommended design strategy would be to use area units at cost. each successive stage of sampling until it is no longer feasible to define area units of the desired size. Only at that point, i.e., as a last resort, should list frame units be used.

Many countries use the same kind of frame unit, list or area, for every sampling unit at a given stage of sampling. While this approach is operationally simple, it does not always make optimum use of existing possibilities for creating area segments. An alternate approach, illustrated in some of the case-studies (see, for example, Botswana) is, at a given stage of sampling, to divide the selected sampling units into area segments varying in size, but with each segment being as small as possible, subject to some minimum number of housing units. When any one of these area segments falls in the sample, it may be treated as a take-all segment if it is small enough; otherwise it will be listed and a sample of listing units selected. With this approach, the advantages of area segments are more fully exploited at the cost of a moderate increase in operational complexity.

When the stage is reached where list-frame units must be used, under most conditions the preference for housing units expressed in the UN <u>Handbook of Household Surveys</u> (1984) and the WFS <u>Manual on Sample Design</u> (1975) is affirmed. Households should be considered for use only if the list frames are to be used for one or more surveys, all of which are to be conducted within a fairly short time period. The practice of using households as units for SSFs in situations where the lists or the samples selected from them will be used on more than one occasion cannot be supported. Perhaps it is a carryover from the usual practice of using households as the basic listing units in population censuses. Even for censuses, however, there can be some advantages, especially in connection with the use of census materials for sampling-frame development, in identifying both housing units and the households associated with them.

CHAPTER V MASTER SAMPLES

All of the countries to which the ll case-studies refer use master samples of some type in their integrated household survey programmes (IHSPs). Indeed, it is unlikely that any country would develop an IHSP design without making some use of the master sample concept, as it has been defined for this technical study. The goal of this chapter is to provide a systematic review of the historical development and current use of master samples, so that IHSP designers will be able to consider alternatives and to choose the kind of master sample, if any, best suited to their needs.

Section A describes the prototype master sample and reviews the definition adopted for this study, using some examples of different types of master samples. Section B examines the advantages, which are substantial, of using master samples and the limitations, which must also be considered when developing a master sample design for an IHSP.

In Chapter III, two distinct designs for IHSPs were identified: design A, a single multi-subject survey conducted on a continuing or periodic basis, and design B, a programme consisting of two or more separate surveys on different topics. Section C of this chapter covers the use of master samples in IHSPs using design A, with examples from the case-studies and discussion of several specific design issues. Section D examines the use of master samples for IHSPs consisting of multiple surveys. Finally, Section E reviews some special issues in the use of master samples in IHSPs.

A. What is a master sample?

1. The first master sample

A 1945 article by King credits the idea for a master sample to Rensis Likert, who was then employed by the U.S. Department of Agriculture (USDA). In 1943, when a plan for a master sample was first developed, the USDA was conducting a large number of farm surveys, using probability samples. Methodological studies had suggested that small area segments, each containing only a few farms, would be the most efficient USUs for these surveys. However, the cost and the time needed to develop frames and select separate samples for each survey were considerable.

Likert therefore proposed that a large master sample be selected, so that subsamples for different studies could be selected from it. He saw two advantages to the use of a master sample. The more obvious advantage and the one that has been more often realized in practice is the reduction of the overall costs of providing samples for multiple surveys or survey rounds. However, Likert also felt that the accumulation of data from surveys on different topics using the same master sample of farms would permit the study of important aspects of farm production, income and living that could not be covered in a single survey.

The design of the master sample based on Likert's concept is described in the case-study for the United States of America. In practice, the U.S. master sample of agriculture was an outstanding success, at least in meeting the first of its two objectives. Fuller (1984) cites estimates that the master sample was used to select 60 to 80 samples per year during its first 10 years. Updated versions of the original master sample are still being used. Although there may have been a few earlier applications of the master sample concept as it is now understood, there can be no doubt that Likert and his colleagues at the USDA, the U.S. Bureau of the Census and Iowa State University's Statistical Laboratory were responsible for clarifying and giving a name to the concept and for convincingly demonstrating its utility.

There is less evidence to indicate whether Likert's second goal, the integration of data from different surveys based on the master sample, was achieved. This feature, which represents a possible advantage of using master samples, is discussed further in section B.

2. Definition and key features of a master sample

A master sample was defined in section E of Chapter II as "a sample from which subsamples can be selected to serve the needs of more than one survey or survey round." The essential elements of the definition are: first, that the master sample must be used for more than one survey or survey round and, second, that the subsamples used not be identical for all of these surveys or survey rounds.

Subsampling from a master sample takes many forms. For example, consider a master sample consisting of a probability sample of census EAs, to be used as a basis for several two-stage samples, in which the USUs are to be housing units. One method of subsampling from the master sample would be to select a new subsample of EAs for each survey (or survey round) and to prepare housing unit listings for the EAs in each subsample, as needed. The subsamples of EAs could be selected independently or by a controlled selection process designed to make the overlap between subsamples as small or as large as desired. If a particular survey were designed to cover only a specific geographic area, say a single state or province, the subsample would be limited to that area and could include either a subsample of the master sample EAs in that area or all of them. If the full set were not large enough, a new sample of EAs could be added to those in the master sample from that area.

Another method of subsampling would be to include all of the master sample EAs in every survey, but to select new samples of housing units from these EAs for each survey. Like the subsamples of EAs, the second-stage samples of housing units could be selected independently for each survey or by a procedure that controlled the amount of overlap between surveys. For this method of subsampling, housing unit listings could be prepared for all of the master sample EAs and be used as the secondary sampling frames for all surveys or survey rounds within a specified time period, say one year, during which updating of the listings is considered unnecessary.

Some combination of the two procedures -- subsampling of the master sample EAs and sampling of housing units in the full set of EAs -- could also be used.

So far, in this illustration, the master sample has been described as a single sample of census EAs, selected without replacement. It would also be possible, however, for the master sample of EAs to consist of several independently selected probability samples of EAs, or replicates. The subsampling from this kind of master sample would be done by assigning one or more of these replicates to each survey or round. Since the replicates are selected independently, it would be possible for some EAs to appear in more than one replicate.

Starting from this specific illustration of one set of master sample designs, it is now possible to identify the main features that distinguish different kinds of master samples used in practice. These key features are shown in Exhibit 5.1 and are discussed below.

The four design features listed in the first part of Exhibit 5.1 are sample design features. These features do not fully describe a particular master sample design, but are those considered to be of greatest importance for master samples. The same features apply, of course, to the design of any sample, whether or not it is to be used as a master sample.

A master sample can be selected in one or more stages. In the example above, the master sample of census EAs could have been selected in a single stage or in two or more stages. For a two-stage design, administrative districts might have been used as PSUs and census EAs as SSUs. If a two-stage design were used, additional methods of subsampling for specific surveys would be available, for example:

- Include a subsample of PSU's and all master sample EAs in those PSUs.
- o Include a subsample of PSUs and a subsample of the master sample EAs in the selected PSUs.

However, if a fixed sample of PSUs (administrative districts) had been selected and the sample EAs or other SSUs were then selected from these PSUs separately for each survey, it would no longer be appropriate to refer to a master sample of census EAs; the master sample would then be

Exhibit 5.1 Key features of a master sample

	Number of stages
	Units at each stage
DESIGN FEATURES	Use of replication: full, partial, none
	Selection probabilities of sampling units
	Secondary sampling frames
	Маря
	Types of surveys:
	o Single multiround
THTEMEN HER	o Multiple surveys
THIENDED OPE2	Methods of subsampling
	Duration and frequency of use

-

a one-stage sample of administrative districts. Thus, a master sample can be characterized both by the number of stages of selection and by the type of unit serving as the USU at the time the master sample is selected, e.g., a two-stage master sample of census EAs. Additional stages of sampling may, of course, be introduced when subsampling for a particular survey.

The use of replication in master samples appears to be rather infrequent in practice, but replication designs have distinct characteristics, so it is important to know whether or not a master sample consists of independently selected replicates. The distinction between full and partial replication (see Exhibit 5.1) applies to multistage master samples. For full replication, the selection of each replicate is independent at every stage. An example of partial replication would be a master sample consisting of a fixed set of PSUs from which two or more fully independent samples of census EAs had been selected.

For proper use of a master sample, selection probabilities of units selected at each stage must be accurately recorded. This information will be needed in order to establish the appropriate selection probabilities each time the master sample is used to provide a subsample for a particular survey and to determine what sample weights to use in developing estimates from these subsamples. In some master sample designs, selection probabilities for the master sample units are determined in a manner that facilitates the selection of self-weighting subsamples, i.e., subsamples for which the overall selection probability of every USU, taking into account the selection probabilities of the master sample units from which they were drawn, is the same.

Auxiliary materials for a master sample are usually prepared as needed for the selection and use of subsamples in specific surveys or survey rounds. The most important auxiliary materials are secondary sampling frames for the master sample USUs. Secondary sampling frames (SSFs), which were discussed in Chapter IV, section E, may consist either of list units, such as housing units or households, or area units. In either case, the cost of preparing and maintaining the SSFs is often substantial. This suggests, first, that they be prepared only when needed and, second, that they be used to select samples for use in more than one survey or survey round, whenever this is compatible with other design objectives.

The development of SSFs consisting of area units requires preparation of detailed maps of the master sample USUs on which segments of the desired size can be delineated. Suitable maps or sketches prepared for use in a population census are often available for at least some of the master sample USUs, but sometimes it may be necessary to prepare new maps specifically for use in connection with the master sample. In addition, whether list or area unit SSFs are used, it may be desirable to have smaller-scale master maps showing the locations of the master sample USUs. Master samples can also be characterized according to their intended uses. It may not be possible to anticipate all ways in which a particular master sample will be used, but a review of expected uses is necessary in order to develop a design well suited to those uses.

A master sample that is to be used for multiple surveys covering different topics and, possibly, different areas of the country will normally require greater flexibility than one which is to be used only for a single multiround survey. Examples of master sample designs for the latter situation are given in section C of this chapter and for the former in section D.

The expected methods of subsampling for specific surveys or survey rounds should also be considered in designing a master sample and in deciding what auxiliary materials will be needed. Will all of the master sample USUs be included in the sample for each survey or only a subsample of them? Should the master sample consist of replicates, one or more of which will be used in each survey or round?

Finally, what is the expected frequency of use of the master sample and for how long will it be used prior to any major revision? Answers to these questions will help to determine the size of the master sample and to decide whether to develop list or area SSFs for the master sample USUs. If the master sample is to be used for more than one year, some updating procedure to reflect significant changes in the distribution of the population will probably be required.

This section has served a dual purpose: to identify the key features that distinguish one master sample design from another and to identify some of the factors that must be considered in choosing a suitable design. As a further aid in selecting from among the many possible designs, the following section examines specific advantages of using master samples, as well as their limitations.

B. Pros and cons of using master samples

It is assumed in this technical study that some type of master sample is an indispensible component of a well-designed IHSP. However, readers should not be asked to accept this argument on faith. Therefore, it is now appropriate to consider the advantages of using master samples and also to discuss their limitations.

1. Advantages

A paper by the United Nations Statistical Office (1983) summarizes the benefits of using a master sample in an IHSP:

The use of common system and arrangements for selecting samples for various household surveys is among the most important instruments for achieving substantive integration as well as operational coordination between surveys. In fact, it is often convenient to select a single sample -- whether of area units, dwellings or households -- which is large enough to permit subsampling from it for a number of surveys conducted over a period of time. The use of such a master sample can be very cost-effective in avoiding duplication of the work involved in compiling the necessary sampling materials and selecting the samples. Consequently, more resources and effort can be devoted to improving sample design (for example through better mapping, segmentation, listing of area units) and the costs can be spread out over many surveys. At the same time, samples for individual surveys can be selected more quickly and economically; operational and substantive linkage between surveys and survey rounds can be better controlled. This arrangement would also facilitate the use of the same sampling materials by different agencies, as well as the accommodation of ad hoc needs.

Clearly, the main benefit to be expected is <u>efficiency</u>, i.e., production of the desired survey results at less cost. Other potential benefits are improvements in the <u>quality</u> of survey results and greater <u>flexibility</u> to respond quickly to needs for data on a variety of topics.

Efficiency in an IHSP results from what economists call economies of scale. Some economies of scale are possible even if samples are selected independently for all surveys. For example, if a questionnaire module is used in more than one survey, the one-time costs of developing it and preparing processing procedures can be spread out over all of the surveys in which it is used.

The more important economies, however, are those that are made possible by the use of master samples. As shown in Exhibit 5.2, there are three components of an IHSP for which the use of a master sample may permit costs to be spread over several surveys or rounds.

The selection of a master sample is usually an office operation, requiring little or no field work. It is a one-time operation and the costs of professional staff services (for sample and system design work), clerical operations and computer running time can be attributed to all surveys for which the master sample is used. The total cost of selecting a master sample is small relative to the other two IHSP components listed in Exhibit 5.2, so the potential savings are only moderate. Nevertheless, savings can be realized at this stage regardless of how the master sample is designed and, if the outputs are carefully planned, additional costs of subsampling from the master sample can be reduced. For example, if different subsets of the master sample USUs are to be used in different surveys or rounds, these subsets can be designated at the time the master sample is selected.

Development of the field staff is a second IHSP activity whose costs can be spread out over multiple surveys. Development costs include the recruitment and selection of interviewers and other field staff, as well

Exhibit 5.2 Economies of scale resulting from use of master samples

IHSP component	Costs allocable to multiple surveys	Type of design for which relevant
	Sample design	
Office selection	Systems development	All designs
oi master sampre	Clerical and computer operations	
	Interviewer selection	Multistage
Field staff development	Interviewer training	designs with
development	Supervision	resident interviewers
Preparation and	Map preparation	Destana ta ubtab
SSFs and auxiliary materials	Listing of housing units or households	SSFs are re-used

as training that is relevant to all surveys, e.g., training in interviewing techniques and administrative procedures. To some extent, these costs can be spread out over surveys that use independent samples, provided the same persons do the field work. However, in some countries or regions, given the costs of travel and the conditions of employment, it is considered more feasible and economical to use an interviewing staff based in the sample PSUs. Under these conditions, relatively large PSUs are used, so that there will be a reasonable work-load in each PSU for one or more interviewers. There is then a clear gain from using a master sample of such PSUs, rather than selecting a new set for each survey, since fewer interviewers will have to be be selected and trained in the former case.

The greatest economies of scale, however, can be realized when the same subsampling frames (SSFs) are used for more than one round. Except in unusual cases where current housing unit listings or very detailed maps are already available for the master sample PSUs, extensive field-work will be needed to establish the necessary SSFs. The cost per survey of this field-work and associated office processing will decrease almost in direct proportion to the number of surveys or rounds in which the same SSFs are used. If the SSFs need some form of updating, the cost should still be substantially less than the cost of preparing SSFs for new PSUs or SSUs.

The effect of re-using SSFs in the same master sample USUs will depend somewhat on the survey objectives. If measurement of change in a multiround survey is an important goal, then full or partial overlap of sample units from round to round will result in smaller sampling errors of estimated change for many items. On the other hand, if the object is to accumulate data to estimate aggregates or averages for a period covering several survey rounds (e.g., four quarterly rounds in a one-year survey), the use of overlapping units may increase sampling errors for some items. It is implicit in these comparisons, however, that the sample sizes are the same for the overlapping and non-overlapping designs. When costs are taken into account, savings from the re-use of SSFs can make it possible to use larger samples for the overlapping design and thereby compensate to some degree, or perhaps even completely, for its loss of efficiency in the estimation of aggregates.

Another possibility, of course, is to use the master sample for multiple surveys on different topics. If this is done in a way that permits some use of the same SSFs for different surveys, there will clearly be some economies of scale. Owing to differences in survey objectives, it may not be feasible to have complete overlap of the master sample USUs included in the different surveys, but even partial overlap can bring significant savings. In addition, if it is desired to integrate data from two or more surveys for analytical purposes, the use of overlapping sample units can be a considerable advantage. Some examples of such substantive integration of data from different surveys are given in a United Nations Statistical Office paper (1983). The realization of benefits from substantive integration of data from different surveys depends on three factors: the extent of sample overlap, the relationship of survey reference periods and the ability to process the linked data.

If two surveys to be linked have full or partial overlap of USUs (i.e., data are to be obtained in both surveys for the same households or persons), the linkage can be used to create a database nearly equivalent to that which could have been created from a single survey covering the same topics. Relationships among variables from both surveys can then be analyzed at the level of the elementary units. If the sample overlap between the two surveys is for PSUs or SSUs, but not for elementary units, analyses of relationships between variables will have to be based on aggregate data. The sample overlap will reduce the sampling errors of estimates based on variables from both surveys, but less powerful analytic techniques will have to be used. Although these considerations tend to favour inclusion of the same household in two or more surveys, such overlap increases the response burden on these households, so it should not be carried to extremes.

Survey linkages for analytical purposes are useful mainly when the data from all surveys cover essentially the same time period. This is especially true when individual households are being linked. Changes in household composition over time make it impractical or difficult to link data for the same households from surveys covering different time periods. Even if linkages can be made, the validity of analyses of specific relationships may be adversely affected by changes in household composition.

Finally, and perhaps of most importance, the creation of a database by linking records from two or more surveys requires considerable sophistication in systems design and data processing: it is substantially more difficult than creation of a usable database from a single survey. Thus, although substantive integration of data from two or more surveys, made feasible through the use of a master sample, may seem an attractive possibility, it should not be oversold as a likely product of an IHSP in its early stages of development.

Use of a master sample creates opportunities for improving the quality of survey data. When master sample USUs are to be used for several surveys or rounds, it may be feasible to use more resources to create the maps and SSFs associated with them. Accurate maps with clearly delineated boundaries can lead to reduction of errors in coverage of area segments assigned for housing unit or household listings. More thorough quality control procedures can be adopted to ensure the quality of such listings.

Some master sample designs facilitate the use of a permanent field staff, so that the same supervisors and interviewers can be used for all or most surveys. The quality of interviewing does not automatically improve with experience (cf, Rustemeyer, 1977), but normally will, given adequate attention to training, supervision and quality control. For all practical purposes, a permanent field staff, at least down to the supervisory level, can be regarded as a necessary although not sufficient condition for acceptable performance of field work.

Needs for surveys often arise unexpectedly. An economic crisis or a natural catastrophe may generate data needs that are not covered by even the most carefully planned IHSP. Other government agencies may request help in conducting surveys designed to meet their special requirements. A well-designed master sample can be used to provide quickly the sample units and associated maps and SSFs needed for such ad hoc surveys. Such flexibility will be possible, of course, only if the initial master-sample design takes into account these unpredictable requirements.

2. Limitations

Master samples have limitations, both in their flexibility to meet the design requirements for surveys on widely varied topics and in the length of time for which they can be used without major revision or redesign. In addition, there are some possible negative effects on quality that should be considered. None of these features leads to an outright rejection of the use of master samples, but awareness of them should guide the design of master samples for use in IHSPs.

To what extent can a master sample designed for general purpose national household surveys be used for surveys with different data requirements? Special data requirements may include:

	 Data for selected political divisions, e.g., states or provinces.
SPECIAL DATA	- Data for unevenly distributed subgroups, e.g., specified ethnic groups.
REQUIREMENTS	- Data for non-household units, e.g., institution- alized population, farms, businesses.

- Data for rare items such as disability or persons with higher education.

One option, of course, is to design completely independent samples for surveys that have these special requirements. Under this option, the sample design for each survey can be tailored to the specific data requirements and, considered only in the context of that single survey, optimized. In the context of a programme of surveys, however, a design that uses a master sample, sometimes with appropriate supplementation, may be a better solution.

To design a general purpose household survey for a province, for example, the solution can be straightforward: use master sample USUs to the extent that such use is compatible with the regular survey programme and select additional units, as needed, from the master sampling frame. Where feasible, use existing field staff and SSFs prepared for master sample USUs.

A similar approach could be used for subgroups of the general population that are not evenly distributed throughout the country. Suppose, for example, that one wishes to survey members of an ethnic group that is concentrated in two or three provinces, with only a scattering in the remainder of the country. A possible design would be to supplement the master sample as needed in provinces where the ethnic group is concentrated, and to rely entirely on the master sample, with a small subsampling fraction, in the remaining provinces.

Although this technical study is directed at household surveys, a master sample designed primarily for household surveys may also be considered for use in surveys of other units, such as farms, nonfarm businesses and persons living in institutions. In most developing countries, a large proportion of the farms and nonfarm businesses (especially in the retail and service sectors) are directly associated with households on a one-to-one basis. Furthermore, in many rural areas, a large proportion of households operate farms. Because of these associations, national multi-purpose household samples, with suitable adjustments, can often provide adequate and efficient samples of farms and nonfarm businesses.

Two caveats must be attached to this general statement. First, there are likely to be large units such as state farms, plantations, manufacturing plants and corporate enterprises in general, that cannot be reached efficiently through a master sample designed for household surveys. To cover these units, a dual-frame approach can be used. The large units can be sampled from a separately developed list; all other units can be reached through the household sample. If the master sample consists of large PSUs served by a permanent field staff, it may be more efficient to select list sample units, except for the extremely large ones, only in the master sample PSUs.

The other caveat relates to agricultural surveys. In surveys whose main purpose is estimation of crop areas and production, several countries use area frames that have land parcels as USUs and do not involve households at any stage of sampling. If up-to-date, accurate land records are available this is clearly the preferred approach. If the country's master sample design uses large PSUs with permanent field staff based in the sample PSUs, there might still be benefits from overlap in the household and agricultural survey samples at the PSU level; however, in many countries the surveys are completely independent, with the field-work being performed by two different groups.

There is no direct link between the institutional population and the universe or a sample of regular households. If coverage of the

institutional population is wanted, it will be necessary to prepare a list of institutions to use as a sampling frame. However, depending on the master sample design, it might be feasible and efficient to prepare lists of smaller institutions only for the master sample PSUs. As was suggested for farms and nonfarm businesses, the sample of institutions, except for extremely large ones, could be restricted to the sample PSUs and the field-work carried out by the regular household survey interviewers.

Coverage of rare populations, such as blind persons or persons with higher education, requires somewhat different approaches. There are two ways in which effective use can be made of master-sample techniques to get an adequate sample of such persons. The first method is to identify such persons (or households) in multiple rounds of a continuing household survey, as part of the regular survey interviews. This method is only effective, of course, if there is complete or partial rotation of the sample households from one round to the next. Depending on the specific data requirements for the rare population, the data might be collected as part of the regular household survey interview or at a subsequent time.

The second method applies to any designs that use listing and sampling of households, whether or not they involve the use of a master sample. As part of the listing operation, screening questions can be asked to identify households with one or more members of the rare target population. All or a sample of these households would be interviewed at the time of listing or later on to obtain the relevant information.

The first of these two methods has the advantage that all of the standard household survey data would be collected for all sample households containing members of the rare population (as well as for other sample households). For the second method, on the other hand, this information would not be available as a matter of course for all households with members of the rare population and, if needed, would have to be collected in the course of interviewing those households.

Another limitation of master samples is that in certain respects they deteriorate over time. Changes occur that affect the definitions of the master sample units and the measures of size that were used in their selection. These changes also affect the SSFs prepared for sampling from the master sample USUs. SSFs consisting of household or housing unit listings are especially vulnerable to changes.

Some kinds of changes can cause coverage bias as, for example, when no steps are taken to add new housing units to list SSFs for the master sample units. Changes in measures of size of master sample units or units used in area SSFs are likely to lead to increased sampling errors. An extreme but not unheard of example would be construction of a housing project with 50 housing units in a segment that previously contained only five units.
There are various ways of coping with such changes; these are discussed in section C of this chapter. However, sooner or later there is a time when these updating and adjustment procedures are no longer cost-effective, and an entirely new master sample is needed. The length of time for which a master sample can be used depends on the types of units used for the master sample and the associated SSFs, and on the extent and distribution of changes in the survey target populations. Area units are generally less affected than list units. The initial costs of preparing SSFs with area units may be higher, but these additional costs can be recovered if the SSFs can be used longer without updating.

There are some ways in which the use of a master sample might adversely affect the quality of survey results. One conjecture is that the master sample and subsamples drawn from it might become "unrepresentative" because of conscious decisions by government agencies to concentrate economic development and social service programmes in the sample areas, resulting in estimates of economic and social indicators that provide an overly optimistic picture of the country's living standards and rate of development. This is a possibly extreme example of a more general phenomenon called conditioning, i.e., the effects of the measurement process on the units being measured. There 18 convincing evidence of conditioning from multiround surveys, on various topics, that use partially rotating sample designs. Estimates based on samples of households that have been in the survey in prior rounds frequently differ significantly from estimates based on households in the survey for the first time (Bailar, 1975).

Conditioning through repeated interviews can affect survey results in many different ways. Respondents may feel overburdened and become less inclined to give accurate responses or to respond to the survey at all. Interviewers may tend to copy responses from earlier survey rounds under the assumption that no change has occurred. The phenomenon is complex and its causes and effects are not fully understood (United Nations, 1982a).

Although the existence of conditioning effects in panel surveys has been conclusively demonstrated, there is still considerable uncertainty about the magnitude of biases caused by conditioning and whether these biases are more likely to increase or decrease over the series of survey rounds for which units are included in the sample. For the purposes of this discussion, it can be recommended that survey designers, in considering master sample uses that require a panel design, i.e., the inclusion of individual reporting units in two or more rounds, be aware of the possibility of respondent conditioning and do whatever is possible to limit its influence on survey results.

It has also been suggested that updating of SSFs for master sample units (as opposed to independent development of SSFs for new units) can lead to biases. This argument, which would apply primarily to SSFs consisting of housing-unit or household listings, supposes that the creation of entirely new listings for sample areas would, on the average, be subject to smaller listing errors than the updating of existing listings.

This is an empirical question. Whether the hypothesis of greater listing error for the updated SSF is true or not would depend on the quality of the prior listings and on the care exercised in supervision and control of the listing operations. Even if the update approach should lead to moderately higher listing errors, the lower cost of this approach might make its use preferable when the criteria of total survey design are applied.

Finally, it can be argued that master sample designs tend to be more complex than ad hoc designs developed separately for each survey. By implication, the likelihood of error in the development and execution of a master sample design is greater than it would be for a single survey.

This is no doubt a correct assertion. Nevertheless, many countries have used master samples in their household survey programmes, with substantial benefits in terms of efficiency, quality and flexibility. Others wishing to do the same should be aware that the process requires a certain level of sophistication in sampling and survey design and, if necessary, should seek qualified assistance from outside sources to help develop the initial design and monitor its execution.

3. Summary

The main points that have been made in this section are:

- a. The use of master samples can improve efficiency in several different ways (see Exhibit 5.2). The greatest gains come from the use, in more than one survey or survey round, of secondary sampling frames developed for sampling from master sample USUs.
- b. Some of the savings resulting from use of a master-sample design can be applied to quality improvements through the development of a well-trained field staff and of better quality maps and subsampling frames.
- c. Master samples can provide flexibility in responding quickly to new data needs.
- d. Master samples have some potential limitations that need to be considered in deciding what kind of master sample design to use. These limitations include:

- (i) Some compromises in sampling efficiency may be necessary to accommodate widely varying data requirements.
- (11) Like a master sampling frame, a master sample cannot be used indefinitely. Minor adjustments are needed from time to time. Less frequently, probably after each population census, a major redesign is necessary.
- (111) When sample units are re-used, especially at the household level, there is a possibility of biases resulting from conditioning effects.
- (iv) Designs for master samples used in connection with IHSPs tend to be somewhat more complex than designs for ad hoc surveys.

A suitable master sample can be a desirable feature of an IHSP design, provided the benefits outweigh the limitations. A review of the casestudies shows that many survey organizations have adopted the master sample concept in their designs. In choosing a particular application of the master sample idea, however, survey designers should be aware of and prepared to deal with the limitations described in this section.

The next two sections of this chapter draw on the case-studies to show how several developing countries are applying the master sample concept. Some illustrations from more developed countries are also included.

C. Use of master sample principles in multiround surveys

As explained in Chapter III, some countries have IHSPs consisting of a single multiround, multi-purpose survey, some conduct multiple surveys on different topics, and some combine the two approaches. This section discusses the application of master sample principles in a single multiround survey. First, some examples from the case-studies are presented and their key features compared. Then specific design issues are discussed, with illustrations taken from case-studies and other sources.

1. Some examples

Three developing country examples, selected from the case-studies for Jordan, Saudi Arabia and Thailand, illustrate the extent of diversity between countries in how they apply master sample principles in a multiround survey. Each design is described briefly (for further detail, see the case-studies in Appendix A); then key features of the three designs are compared.

<u>Jordan</u> - The proposed master sample consists of 21 independently selected samples (replicates), each consisting of 35 urban and 15 rural census EAs (called "blocks" in the case-study) or groups of adjacent EAs. Urban EAs or EA groups were to be selected in one stage and rural EAs or EA groups in two stages, with localities serving as PSUs.

Each EA or EA group was to be selected with probability proportionate to an assigned integer measure of size, and structures or households were to be selected from each sample EA or EA group using a sampling fraction equal to the reciprocal of its measure of size. SSFs would be updated census listings of housing units. The master sample was designed for use in a multi-purpose household survey over a five-year period. Two survey rounds were planned for the first two years and four rounds per year thereafter. One or more of the 21 replicates would be used for each round. No specific proposal was given for the selection of replicates for each round, but a set of general guidelines included in the proposal can be interpreted as recommending partial replacement of replicates from one round to the next.

<u>Saudi Arabia</u> - The master sample was a self-weighting sample of approximately 300 segments in 21 PSUs. The PSUs were emirates, an administrative subdivision, and each of the area segments was expected to contain from 100 to 200 households.

For each master sample segment, a listing of structures, housing units and households was prepared. Each household in a

sample segment was assigned by a random procedure to one of eight panels.

The master sample was designed for use in a multi-purpose household survey over a five-year period. Four of the eight panels were included in the sample for the first year. In each subsequent year, one of the four panels from the prior year was to be replaced by a new one.

<u>Thailand</u> - The proposed master sample was to be a one-stage sample of blocks (subdivisions of census EAs) in urban areas and villages in rural areas. Within defined strata, units would be selected with probability proportionate to the most recent available household counts.

For each master sample block and village, a listing of housing units was to be prepared shortly before the first scheduled interviewing in that sample unit. A systematic sample of housing units, 15 in urban areas and 10 in rural areas, would be selected from each listing.

The master sample was to be used in a proposed multi-purpose survey consisting of four survey rounds in a one-year period (after each annual survey, a new sample of blocks and villages was to be selected). The master sample was to be divided into five random subsamples, each consisting of one-fifth of the sample blocks and villages. Subsamples 1 and 2 would be used in the first quarterly round of the survey, subsamples 2 and 3 in the second round, and so forth, resulting in a 50-percent overlap of the samples for successive quarters within a year.

Exhibit 5.3 compares selected features of the master-sample designs used by these three countries in their multiround household surveys. This comparison illustrates a range of options with respect to the duration of use of the master sample, the extent and nature of sample overlap between survey rounds and the intensity of use of SSFs prepared for master sample units.

The designs developed for Jordan and Saudi Arabia use the same master sample for a five-year period. Thailand's master sample, in contrast, is used for only four survey rounds over a one-year period. Economies of scale (see section B of this chapter) increase in direct proportion to the number of rounds for which the master sample is used. The main reason that Thailand selects a new master sample each year is that villages, which are the PSUs for the rural strata, have been increasing in number at an annual rate of about 4 percent in recent years (Table 4.1). To use a master sample of villages for a longer period would require an elaborate updating procedure to ensure the representation of new villages and to revise the definitions and SSFs for existing villages directly affected by the creation of new ones.

	Country			
Feature	Jordan	Saudi Arabia	Thailand 1 year	
Duration of use of master sample	5 years	5 years		
Number of survey rounds per year	Years 1 and 2 - 2 Years 3 to 5 - 4	4	4	
Rate of sample replacement	Not specified	25% at end of each year, none within year	50% each quarter	
Number of rounds housing units remain in sample	Not specified	Minimum - 4 Maximum - 16	Minimum - 1 Maximum - 2	
Type of units used for replace- ment	Independent replicates	Different panels of housing units in a fixed sample of segments	New PSUs (blocks and villages)	
Proportion of listed housing units sampled	50% (estimated)	100%	10% (estimated)	

.

Exhibit 5.3 Key features of master samples used in multiround surveys: Jordan, Saudi Arabia and Thailand

The rate at which sample units are replaced after each round (the complement of the percent overlap between rounds) can vary from 0 to 100 percent. The replacement rate proposed for Thailand -- 50 percent after rounds 1, 2 and 3 each year and 100 percent after round 4 -- was much greater than the rate planned for Saudi Arabia -- 0 percent after rounds 1, 2 and 3 and 25 percent after round 4. The plan for Jordan did not specify a replacement rate. The choice of a replacement rate depends on part on a judgment about the relative importance of estimates of aggregates and estimates of change. A high replacement rate leads to more reliable estimates of aggregates based on data from two or more rounds and a low rate leads to more reliable estimates of change. Respondent burden must also be considered: interviewing the same respondents over too long a period may produce substantial biases due to non-cooperation and other conditioning effects. In the Saudi Arabia design, a sample household could be interviewed as many as 16 times over four-year period, whereas in Thailand, the maximum number of a interviews is two in successive quarters. A still different pattern is used in the United States Current Population Survey, where sample households are interviewed in four successive months, then leave the sample and return to it for the same four calendar months after an eight-month absence. In practice, then, replacement rates varv considerably, depending on the survey designers' judgments about the relative importance of different kinds of survey estimates and the possible adverse affects of excessive response burden.

Also of importance is the choice of the kinds of sample units used for replacement. In multistage sample designs, the replacement units can be anything from different USUs, selected from the same sample units at the next higher level, to different PSUs. Both alternatives appear in the three examples. Replacements for Jordan and Thailand consist of new PSUs. The Jordan design uses full replication: the selection of PSUs, segments and households in each replicate is independent of all other selections. The PSUs used for replacement in Thailand are random subgroups of PSUs (urban blocks and villages) from the master sample of PSUs.

In Saudi Arabia, on the other hand, the replacement units are different households in a fixed set of sample segments. At the start of the period during which the master sample is to be used, household listings are prepared for all of the sample segments. All of the households in each segment are assigned to one of eight random subgroups (called panels). Four of the eight panels are used in the survey during the first year. At the start of each subsequent year one of the old panels is to be replaced by a new one. What are the consequences of these substantially different replacement systems? First consider their effect of the reliability of estimates of year-to-year change. For this purpose, assume that Jordan will use four replicates during a survey year and will replace one of these each year. Exhibit 5.4 shows the nature of the year-to-year overlap for these three designs.

Country	Same housing units	Different housing units, same SSUs	No overlap	
Jordan	75	0	25	
Saudi Arabia	75	25	0	
Thailand	0	0	100	

Exhibit 5.4 Year-to-year overlap (percent) for three master-sample designs

From the figure it is evident that, other things being equal, the Saudi Arabia design would produce the most reliable estimates of year-to-year change, but that estimates for Jordan would be almost as reliable. The Thailand design does nothing to improve estates of year-to-year change.

Second, how does the choice of replacement unit affect the cost of preparing SSFs? In Jordan, whenever a replicate is replaced, new housing unit listings must be prepared for each SSU included in the new replicate. In Thailand, new listings are required each quarter for the new sample blocks and villages. In Saudi Arabia, on the other hand, replacement panels consist of new housing units from the same sample segments, so that no new listings would be necessary at any time during the five-year period for which the master sample is to be used. However, unless some procedure were established for periodic updating of the listings, serious coverage biases could develop after the initial rounds of the survey.

Another way of making this kind of comparison is to look at the intensity of use of the list SSFs. As shown in Exhibit 5.3, only about 10 percent of the listed housing units in Thailand are ever included in the sample. In Jordan, where the ultimate cluster size is larger and the average size of listing units is smaller, roughly half of the listed units are included in the sample. In Saudi Arabia, however, all of the listed units would be included in sample for at least one year during the five-year period for which the master sample is to be used.

2. Discussion of design issues

The preceding comparison of master-sample designs for three developing countries demonstrated the diversity of designs in

current use. Now each of the principal issues that arises in designing a master sample for use in a multiround survey will be reviewed. It is not the object of this review to specify exactly which of the available alternatives should be preferred in every possible set of circumstances. The purpose of the review is to identify the considerations that should guide survey designers in making these important design decisions. Readers faced with these apparently difficult decisions may take comfort from the fact that any one of a wide variety of designs can meet a country's requirements adequately, even though it does not necessarily achieve the best possible results according to the criteria of total survey design. Time and experience will permit fine tuning of the design to bring it closer to the optimum.

a. <u>Overlap between rounds</u>. Most master sample designs use some overlap between survey rounds. As shown by the examples in subsection C,l of this chapter, there are two kinds of overlap. The most direct type is overlap between USUs, i.e., inclusion of the same housing units, households or compact area segments in the sample for more than one round. The other kind of overlap retains sample SSUs or PSUs in the sample for more than one round but replaces the sample USUs. In the examples for Jordan, Saudi Arabia and Thailand, all three countries use direct overlap of USUs to some degree. The Saudi Arabia design also uses overlap of higher level units throughout the entire period of use of the master sample.

There are also options for the duration of the overlap (number of rounds) and for its timing. For most designs, the overlap lasts for a specified number of rounds. However, it is also possible to rotate units out of the sample and bring them back later, as is done in the United States Current Population Survey.

The principal advantages and disadvantages of overlap have been discussed in connection with the examples. They can be summarized as follows:

ADVANTAGES o Economies of scale

- o Reduction of sampling error for estimates of change
- DISADVANTAGES o Negative effects on quality from excessive response burden and conditioning of respondents
 - o Increase in sampling error for estimates made by aggregating data over survey rounds

For the most part, both the advantages and adverse effects are accentuated by using direct overlap of USUs and by increasing the duration of the overlap. In practice, this leads to compromise solutions such as partial overlap between rounds or full overlap between rounds within a twelve-month period and partial or no overlap between annual survey periods. Many designs use direct overlap for relatively short periods and indirect overlap for longer periods. The United States Current Population Survey, for example, uses a fixed set of PSUs for a ten-year period with partial replacement of USUs for each monthly round.

The extent of a country's need or desire for sub-national estimates can be an important consideration in determining the amount of overlap. To produce reliable estimates for a large number of political or geographic subdivisions requires not only a large sample of USUs, but also a sufficient number of sample PSUs and SSUs for each subdivision. Since it is usually not feasible to maintain a permanent field staff that can handle a sample of this size in a short-duration survey round, the preferred solution is to spread the sample over several rounds, usually within a single year, and to aggregate the data for each subdivision over these survey rounds. For this kind of design, which is exemplified by India's National Sample Survey (see case-study), overlap is clearly a disadvantage.

It might be argued that direct overlap of USUs is necessary for topics that require two or more interviews from sample persons or households in order to obtain accurate data for a specified reference period. Such a longitudinal or panel approach is often used for topics like expenditures, income or vital events. Data collected at each interview provide a benchmark or bound to facilitate accurate reporting of subsequent events or transactions at the next interview. Such panel surveys, however, require special treatment of changes in the composition of sample households. For this reason, they do not fit in easily with multi-purpose multiround surveys in which the direct overlap of USUs is not used for longitudinal analyses. Therefore, true longitudinal surveys are usually designed as separate surveys. Thev can, of course, use the same master sample as the multi-purpose multiround surveys.

b. <u>Stages of sampling</u>. How many stages of sampling should be used in the selection of a master sample? What units should be used at each stage?

To examine these questions, it will be useful to return to the examples for Jordan, Saudi Arabia and Thailand to look at differences in the number of stages and types of master-sample units and the reasons for these differences. Exhibit 5.5 shows the relevant features of these three master-sample designs.

Looking at the number of stages in the master sample, one sees that only a single stage of sampling is used in Thailand and in the urban sector in Jordan. Jordan's design uses two stages in the rural sector and Saudi Arabia uses three stages everywhere. Saudi Arabia's proposed design is unusual in that the final stage of selection of the master sample requires listing of households in a sample of area segments and

Country and	Type of sampling unit			
sector	Stage 1	Stage 2	Stage 3	
Jordan				
Urban	Census EAs (or groups)	NA	NA	
Rural	Localities	Census EAs (or groups)	NA	
Saudi Arabia (all sectors)	Emirates	Area segments	Random sub- groups of households (panels)	
Thailand Urban	Blocks within census EAs	NA	NA	
Rural	Villages	NA	NA	

Exhibit 5.5 Stages of sampling for three master-sample designs

random assignments of the listed households to subgroups which would then be introduced at various times during the five-year period of use of the master sample.

For Jordan and Thailand, in contrast, listings would be prepared for master sample USUs only at times when they were about to be used. Hence, the samples of housing units from these listings are not considered to be part of the master sample. This latter approach is preferable: as pointed out earlier, housing unit listings deteriorate over time, causing coverage problems if the listings are used over too long a period. Furthermore, preparation of listings for all master sample PSUs or SSUs at one time might place an unduly heavy burden on the field staff.

Decisions on whether to use one or more stages of sampling to select a master sample depend on several interrelated factors: population density, the ease or difficulty of travel to sample locations, the kinds of sampling frame materials available and the nature of the field organization. To a considerable extent, these factors are the same as

NA - Not applicable, sampling does not extend to this stage.

those that determine the optimum number of stages in the design for an <u>ad hoc</u> survey. An important difference, however, is that for a master sample it is feasible to use area sampling for smaller areas because the cost of preparing area frames for sample PSUs can be spread over several survey rounds. Thus, in Saudi Arabia, it was possible to divide each of the sample PSUs into area segments, each containing from 100 to 200 households.

In Thailand it was feasible to use a one-stage master sample because population density is high and travel is relatively easy in most parts of the country. The field staff are based in the capitals of the country's 72 provinces (changwads). For the urban areas, detailed block maps were available, so areas smaller than census EAs could be used as PSUs.

Few, if any, of these same conditions existed in Saudi Arabia or the rural areas of Jordan, so multistage master sample designs were more appropriate.

The considerations that determine what kinds of units to use at each stage are essentially the same as those that influence the choice of units for master sampling frames. One should look for units that are well defined, likely to remain stable over time and for which good quality maps and up-to-date measures of size are available. These requirements are discussed in detail in section C of chapter IV.

c. <u>Use of self-weighting samples</u>. A self-weighting sample is one for which all USUs have the same overall probability of selection, taking into account all stages of sampling. The self-weighting feature of a sample for a national household survey may apply to the entire sample or it may apply only within areas for which separate estimates are to be made, e.g., states, provinces or the urban and rural sectors.

When a master sample is used with subsampling for each survey round, the self-weighting property can be obtained in one of two ways. The first is to design the master sample itself as a self-weighting sample. If this is done, the subsampling must be carried out in a way that preserves the self-weighting property. Suppose, for example, that the master sample were an equal probability sample of census EAs and that the sample for each survey round required a sample of housing units selected from a subsample of these EAs. To make the final sample of housing units self-weighting, the product of the selection probability of EAs from the master sample and the selection probability of housing units in the selected EAs would have to be the same for all sample housing units. For example, if the overall subsampling rate from the master sample was to be 1 in 20, one could select EAs and housing units as follows:

Sampling	Sampling		
fraction for	fraction for HUs in selected		
master			
sample EAs	EAS		
Take all	1 in 20		
1 in 2	1 in 10		
1 in 5	1 in 4		
1 in 20	Take all		

One could use one of these patterns for the entire sample or different patterns for different strata.

The second approach is to design a master sample that is not self-weighting, but to subsample in a way that makes the sample for each survey round self-weighting. For example, suppose a master sample of census EAs had been selected with probability proportionate to size, with selection probabilities I/M_1 , where:

M₁ = measure of size of the ith EA I = the overall sampling interval

The subsample could be made self-weighting by selecting a subsample of master sample EAs and housing units within these EAs with overall probabilities C/M_1 , where C is a constant equal to or less than all M_1 . The product of these probabilities:

 $\frac{M_1}{I} \times \frac{C}{M_1} = \frac{C}{I}$

is a constant, so that the overall sample is self-weighting.

The first of these two approaches is illustrated by the design for Saudi Arabia. Each of the eight panels consisting of households in a fixed set of area segments is a self-weighting sample. When a subset of these panels is selected for inclusion in a survey round, this is equivalent to subsampling at a constant rate from a sample of area segments which is also self-weighting.

The Jordan design illustrates the second method in a somewhat more complex way. Subsampling from the master sample consists of two steps: selecting one or more of the 21 replicates (samples of census EAs or EA groups), which is equivalent to subsampling at a constant rate, and then subsampling from listings for the EA groups in these replicates at a rate that makes the overall sample self-weighting. All EA groups were to be selected with probability proportionate to integer measures of size, so that subsampling of households could be done using these integer measures as sampling intervals (this would be equivalent to making C = 1 in the preceding general formulation).

The master sample for Thailand could have been used to produce a self-weighting sample for each of the publication areas, but was not. Master sample blocks and villages were to be selected with probability proportionate to size (the latest available count of households). The sample blocks and villages were to be allocated to five random subgroups and two of these were to be used in each survey round. A fixed number of housing units, 15 in urban strata and 10 in rural strata, were to be selected from each sample block and village. This process would produce a sample for each publication area that was only roughly self-weighting, with the extent of departure depending on the extent of changes in the sizes of blocks and villages.

The main argument for using a self-weighting sample is that its use simplifies processing of the data: a single, uniform weight suffices to produce estimates of aggregates and no weights are needed to estimate ratios such as means, percents and proportions. In addition, a self-weighting sample usually works out to be close to the optimum design for a multi-purpose survey: it is not necessarily optimum for each variable, but better on the average than most alternatives.

One disadvantage is that efficient designs that are self-weighting tend to be a bit more complex than those that are not. A second disadvantage is that use of a self-weighting sample usually makes it more difficult to control precisely the size of ultimate clusters and, hence, interviewer workloads. The use of fixed size ultimate clusters was preferred in Thailand for just this reason, even though it meant that the sample was only approximately self-weighting.

In deciding whether to use a self-weighting design, the advantages are to be balanced against the disadvantages. The use of multiple weights is not a big problem with today's computers and may be necessary, even with a self-weighting sample, to adjust for non-response or for other reasons.

d. <u>Exhaustion of sampling units</u>. Normal practice in multiround household surveys is to replace all or part of the sample from one round to the next or at least from one year to the next. Rotation schemes vary: some designs call for replacement by new USUs in the same SSUs, some call for inclusion of new SSUs in a fixed set of PSUs and some, e.g., Jordan, call for introduction of new replicates, each consisting of an independently selected sample of PSUs, SSUs and USUs. Whatever scheme is used, it is possible, in time, to exhaust one or more of the sample units from which the replacement units at the next level are being selected, i.e., there will be no units left that have not already been included in the sample.

A specific example may help to clarify the concept of exhaustion of sample units. Consider the following design for a survey with quarterly rounds:

- A master sample of census EAs is selected with probability proportionate to size. The measure of size, M₁, for the ith EA is its population census count of households divided by 10 and rounded up to the next integer.
- o Each master sample EA is divided into M₁ area segments.
- o For the first quarterly survey round, one segment is selected at random in each master sample EA.
- o For the second quarterly round, new sample segments are selected in one-fourth of the sample EAs; these segments replace the initial sample in those EAs.
- o For the third round, new segments are selected in a different one-fourth of the EAs, and so on, throughout the period of use of the master sample.

If the master sample is to be used for a five-year period, any EA with a measure of size of five or less is subject to exhaustion, i.e., there will be no unused segments left for a rotation scheduled after all five segments have been used.

One possibility, which is unbiased with respect to sample selection, would be to bring previously-used units back into the sample after the next higher-level unit has been exhausted. In the above example, in an EA with only three segments, the three segments would be used in sequence -- 1, 2, 3, 1, 2, 3, etc. -- so that each segment would return to the sample after an interval of eight rounds. However, re-use of sample USUs is sometimes deemed to be unacceptable, because it places an extra burden on respondents in small EAs.

What are the alternatives? Probably the simpler solution, and the one much more commonly used in the case-studies is to form sampling units large enough so that they cannot be exhausted by the planned use of the master sample. In the above illustration, this could be done by combining census EAs before selecting the master sample. Each census EA having a measure of five or less would be combined with one or more similar (and usually adjacent) EAs to form an EA group having a measure of six or more. The PSUs for the master sample selection would be individual EAs and groups of EAs, all having measures of size of six or more.

This simple solution has one disadvantage: it increases the amount of work necessary to divide master-sample PSUs (EAs and EA groups) into segments. It requires division of all EAs in a group into segments, whereas if larger EAs in the group (those with measures of size of six or more) had been selected separately, it would not have been necessary to segment other EAs in the group. The problem can be overcome to some extent by only combining EAs with small measures of size; however, this may not always be feasible or may conflict with other design requirements.

A second solution is to form the EA groups prior to sample selection and to determine the replacement sequence within the EA group by a random process. In the example, suppose we combine EA-A with $M_1=3$ and EA-B with $M_1=12$ to form a PSU. The hypothetical segments in the two EAs are labelled as follows:

EA-A EA-B 1,2,3 4,5,6,7,8,9,10,11,12,13,14,15

It is known that six segments will be needed during the life of the master sample. To determine, the replacement pattern, we select a random start between 1 and 15. Whenever a replacement segment is needed, the one with the next higher number will be selected. If segment 15 is reached, the next replacement will be segment 1, and so on. With this scheme, for any random start between 4 and 10, all of the segments needed will be in EA-B; therefore it will not be necessary to divide EA-A into segments. For any other random start, both EAs will have to be segmented.

Some readers may wonder why one should not, in this example, select one of the two EAs with probability proportionate to size and exhaust the segments in that EA before proceeding to the other one. This procedure would result in only 3 chances in 15 of having to segment both EAs, as opposed to 8 chances in 15 under the recommended procedure. The answer is that the alternative procedure would be biased, since there would be a tendency, during the life of the master sample, to distort the distribution of sample segments by size of EA: over time, more and more of the sample segments would be selected from the larger EAs.

Within the general framework used in the example, more complex replacement schemes can be developed to minimize the total amount of effort needed to prepare SSFs for master-sample USUs. The case-study for Australia provides one example of such a scheme. Another example is the scheme used in the United States Current Population Survey to replace exhausted PSUs (see U.S. Census Bureau, 1977, Chapter III and Appendix M).

In summary, survey designers should consider the possibility that some master-sample PSUs or SSUs might be exhausted during the life of the master-sample. The problem can easily be avoided by forming larger sampling units prior to selection. More complex procedures that minimize the total amount of work needed to prepare SSFs are also available.

e. <u>Duration of use</u>. How long should a master sample be used before being replaced by a new one? It has been observed in Chapter IV that master sampling frames normally are fully updated at the time of each population and housing census. A full update of a master sampling frame usually implies the need to select a new master sample from the updated frame. At most, for multistage designs using large PSUs, some steps can be taken to maximize the overlap in sample PSUs between the old and new master samples (for a discussion of appropriate techniques, see Keyfitz, 1951).

Thus, for a continuing multi-purpose household survey, a new master sample should clearly be selected each time the master sampling frame is fully updated. But what about the period between frame updates, normally five or 10 years? Should one master sample be used for the entire period or should the initial one be replaced at regular intervals?

This question comes up because of changes in the definitions and characteristics of units in the master sampling frame. Two kinds of changes are particularly important. Boundaries of master sample units may change. This is most likely to occur for sampling unit boundaries that are boundaries of administrative divisions or subdivisions and do not correspond to any physical features, but it can also happen occasionally even when physical features are used as boundaries. Such changes usually require some kind of adjustment in the definitions of the sample units affected, and care must be taken to avoid biases in making the adjustments.

The other kind of change that is critical to the efficiency of the master sample design is changes in the measures of size of the master sampling frame units. Such changes affect sample design efficiency whether or not the units affected are in the master sample. If growth of master sampling frame units were more or less uniform, there would be no problem. This is not the normal pattern of growth, however. Growth tends to be concentrated in certain areas, e.g., the outskirts of cities, near to new highways and in newly opened agricultural lands. Even a small segment within a city block can multiply in size many times due to the construction of a new high-rise apartment building or a public housing development. Such uneven growth patterns increase the variance between units in the MSF and, consequently, the variance of estimates based on samples of these units.

The question of how to update an MSF to reflect such changes was discussed in subsection D,5 of Chapter IV. It was pointed out there that the appropriate method of updating an MSF would depend largely on how it was to be used to provide samples for an IHSP. More specifically, two strategies were identified for the design of a continuing multi-purpose household survey: the sample replacement strategy and the sample revision strategy.

Under the sample replacement strategy, all readily available information on changes is incorporated into the MSF continuously or at frequent intervals and an entirely new master sample is selected periodically from the updated MSF. This strategy is exemplified by the National Sample Survey of India: a new master sample of PSUs (census blocks and villages) is selected each year.

The sample revision strategy retains the same master sample for a longer period, relying on special adjustments to compensate for the effects of changes in the frame and sample units. Various special adjustment procedures can be used. One that is relatively well known is to establish a special "new construction stratum," i.e., a set of areas where large amounts of new construction are known to have occurred subsequent to the development of the MSF, and to select a supplementary sample to represent that stratum. The composition of the new construction stratum can be updated periodically, based on information from official records or from field inspection.

The sample revision strategy is illustrated by the case-study for Australia. A master sample of census EAs (called collection districts in Australia) is maintained for the entire five-year intercensal period. Data on building approvals (permits) are used to update the MSF and to revise the master sample twice a year in areas for which growth has been concentrated in certain EAs. The revision procedure is relatively complex. In strata that meet defined growth criteria, a supplemental sample of PSUs is selected. In strata that do not meet these growth criteria, the existing sample PSUs are reviewed and additional SSUs are selected from those that exhibit unusual growth.

A decision on the duration of use of a master sample implies a decision on whether to use the sample replacement strategy or the sample revision strategy. A master sample can be safely used for perhaps one or two years without any significant adjustments. After that, unless the sample is adjusted to compensate for changes, losses in efficiency and quality may exceed acceptable levels. A choice must be made between the two strategies. There is, possibly, a third option for two-stage master samples with relatively large PSUs: use a fixed sample of PSUs for the entire intercensal period, but replace the master-sample SSUs at shorter intervals.

The sample revision strategy has important advantages. The longer use of the master sample means that the costs of sample selection and preparation of SSFs can be spread over more survey rounds. The longer period of sample overlap permits more reliable estimation of changes over the intercensal period. For designs that use large PSUs, the use of a fixed set of PSUs for a longer period may minimize the need for turnover in the field staff with favourable implications for the quality of field-work.

One advantage of the sample replacement strategy is that each time a new master sample is selected, a sample design that is optimum with respect to the updated MSF can be used. However, the more important advantage of the sample replacement strategy is its simplicity relative to the sample revision strategy. Adjustment procedures needed under the revision strategy to compensate for change can be exceedingly complicated (for an example, see the Australia case-study, reference 2, Chapter 8). Lack of sufficient care in the development and execution of such procedures could lead to substantial sampling biases.

It can be concluded that there are quite substantial advantages to retaining a master sample for the full intercensal period or at least for several years, <u>provided</u> the sampling, data processing and field personnel have the ability to develop and carry out the necessary sample revision procedures. This proviso is meant to be taken seriously; if it is not, the consequences could be unfortunate.

D. <u>Use of master samples for</u> <u>multiple surveys</u>

The previous section covered the use of a master sample as an intermediate frame from which to select sample panels needed for a single multiround survey with sample rotation between rounds. That particular use of master samples may be regarded as a special case of their more general purpose, which is to serve as a subsampling frame for multiple surveys. Substantial economies and other benefits can be realized by using the same master sample as an intermediate sampling frame for several different household surveys. The surveys may be conducted simultaneously, sequentially, or both. Sometimes a master sample designed primarily for the selection of samples for household surveys can also be used to select samples for agricultural or business surveys.

This section will focus on the aspects of master sample design and use that are especially relevant to their use in more than one survey. Aspects already discussed in Section C, e.g., the use of self-weighting samples, will not be taken up again unless there are additional features to be considered.

The section begins with a discussion of four broad objectives that should guide the design of a master sample for use in multiple surveys: economy, durability, flexibility, and simplicity of use. Next, some examples from the case-studies are discussed. Finally, some key design issues are explored in detail.

1. Broad objectives

The design of a master sample for use in multiple surveys should aim for four qualities:

	o Economy
DESIGN	o Durability
OBJECTIVES	o Flexibility
	o Simplicity of use

Earlier in this chapter, it was pointed out that use of a master sample could reduce the cost per survey or survey round for operations such as sample selection, training and supervision of interviewers, and the preparation of secondary sampling frames and associated materials. The greatest economies of scale would result from using the area or list sampling frames prepared for master sample USUs in more than one survey round.

In a multiround survey, the re-use of sampling units of any kind in more than one round can reduce costs and improve estimates of change, but it can also be a disadvantage when the objective is to make estimates of aggregates by accumulating data over two or more rounds. However, this disadvantage does not apply when the master sample is used for two or more separate surveys. Furthermore, the use of overlapping USUs or higher-level sampling units in two surveys covering the same time period creates possibilities for integration of the results from those surveys at the analysis stage.

Response burden and possible conditioning effects still impose limitations on the total amount of overlap between surveys, as do differences in the data requirements for the surveys for which the master sample is to be used. Subject to these limitations, however, a priority objective for the master sample design should be to provide for the re-use of SSFs developed for the master sample USUs, because this is how the greatest economies of scale can be realized.

A master sample can be considered durable if it can be used for a long period with little or no updating. Durability, therefore, is greatest when the master sample PSUs and SSUs are units whose definitions and measures of size are not subject to frequent or extensive changes, i.e., they are stable. The stability of various kinds of sample units was discussed in Chapter IV in connection with frames. The main kinds of sample units are listed in Exhibit 5.6 in order of their relative stability.

Exhibit 5.6 Sample units in order of their relative stability

LEAST STABLE LIST UNITS o Households o Housing Units o Administrative units with no defined boundaries, e.g., villages in Thailand AREA UNITS o Census enumeration areas (EAs) o Small administrative areas o Large administrative areas, e.g., states or provinces MOST STABLE o Areas defined entirely by physical boundaries

If durability is important, the more stable types of units should be preferred to the less stable, both for master sample units (PSUs and SSUs) and for units in the secondary sampling frames created for master sample USUs. If, for example, the master sample is a sample of census EAs, selected in one or more stages, one question that must be decided is whether the secondary sampling frames developed for the master sample EAs will be list frames (housing units or households) or area frames (segments with defined boundaries). The initial cost of developing list frames for census EAs is likely to be less than the cost of dividing them into defined area segments. However, list frames require considerably more updating, so the higher initial cost of segmentation may be justified if the same master sample EAs are to be used over a period of several years.

Flexibility is another important objective for the designer of a master sample. There are two kinds of flexibility that should be aimed at: the ability to accommodate surveys with widely differing data requirements and the ability to provide samples quickly to meet unexpected data requirements.

The geographic areas for which separate estimates are wanted (publication areas) determine, to a considerable extent, how large a master sample is needed and how the sample units should be distributed. Suppose, for example, that some of the surveys included in the long-range plan will require estimates for states or provinces and that others will require only national estimates. For the state-level estimates, census EAs might be appropriate for use as PSUs, and a sample of from 50 to 100 might produce sufficiently reliable estimates.

For the surveys requiring only national estimates, a subsample of the master sample EAs, allocated to the states in proportion to their population, might be used as PSUs. However, in some larger countries, it might be more efficient to use larger areas, e.g., administrative districts, as PSUs for national surveys. In these circumstances, the use of two partly overlapping master samples might be appropriate. The first would be a sample of large PSUs for use in national surveys and the second would be a sample of census EAs in all PSUs. For surveys requiring state estimates the second master sample (of census EAs) would be used. For national surveys the first master sample (of large PSUs) would be used and its second-stage units could be all or some of the census EAs from the second master sample that were located in the sample PSUs.

The foregoing example illustrates only one of many possible master-sample configurations that provide flexibility to meet varying data requirements. There are two general requirements for this kind of flexibility. First, the number of master sample units of any kind must be large enough to meet the maximum anticipated need for that kind of unit. In the foregoing example there should be enough EAs in the second master sample to provide separate estimates for each state or province. Second, in some countries it may be desirable for the master sample system to provide a choice of the kinds of units to be used as PSUs: large units such as administrative districts for surveys using small national samples and smaller units such as census EAs.

What about flexibility to meet unexpected requirements quickly? Here it will be helpful to make a distinction between active and reserve master sample units. Consider a master sample of census EAs. The active sample EAs will be those currently in use for ongoing surveys. For these EAs list or area SSFs will have been prepared and used to select samples for each survey or survey round. The master sample may also contain some EAs that were part of the initial selection but have not yet been used. It may even be that no specific use is planned for some of these reserve EAs during the scheduled period of use of the master sample.

There are two options for using the master sample to provide samples to meet survey requirements that are completely unexpected and cannot be added ("piggybacked") on to existing surveys. Option one is to select a sample of the reserve EAs, prepare SSFs and select the required sample of USUs. Option two is to select a sample of unused USUs, e.g., area segments or housing units, from the active master sample EAs, for which SSFs have already been prepared.

Option two makes possible a quicker response to the new survey requirements, since the SSFs for the sample PSUs (census EAs) are already available. However, there is a cost attached to this quick response. It will only be feasible if the subsampling of EAs for the ongoing surveys is designed so as not to exhaust the SSFs prepared for the sample EAs. If the unexpected needs do not arise, the reserve USUs in the sample EAs may never be used. In a pinch, of course, it would be possible to re-use USUs in the active sample EAs, but this practice might result in exceeding the desired level of response burden, with possible adverse effects on the quality of response.

In spite of the additional costs, option 2 might be preferred. It is probably worth paying some price to be able to respond rapidly to emergency information needs. The statistical office that can do this will serve its country well and will gain respect for its competence.

Simplicity of use and flexibility of the master sample are closely associated. Simplicity of use does not necessarily imply simplicity of design. Indeed, design of a master sample that is flexible and simple to use requires considerably more skill and care than design of a sample for single survey. The master sample design may have to be relatively complex in order to make it easy to use.

The main requirement for simplicity of use of a master sample is that selection of subsamples for individual surveys or survey rounds should be a straightforward process. One way to accomplish this is to select a master sample consisting of several independent replicates. In a multiround survey with partial sample rotation between rounds, the initial sample might consist of two or more replicates, with one or these to be replaced by a new one at the end of each round or survey year. The case-study for Jordan illustrates this method. The amount of overlap between rounds or survey years depends on the number of replicates in the initial sample, e.g., replacement of one out of five replicates would produce an 80-percent overlap. For multiple surveys, different replicates or groups of replicates could be selected for each survey.

Another method of achieving the same kind of simplicity is to select a single master sample and to divide it into several random subgroups which would then be used singly or in groups for individual surveys or survey rounds. In terms of simplicity of subsampling from the master sample, either method, replication or partition of the sample into random subgroups, does essentially the same thing. However, there are some differences between the two methods that are discussed later in this section.

Most master sample designs require sampling of the master sample USUs used in each survey. If it is desired that the overall sample for each survey be self-weighting (see discussion in section C of this chapter), there are two methods that can be used. One is to design the master sample itself to be self-weighting, i.e., all master sample USUs selected with the same overall probability, in which case the survey sample can be made self-weighting by sampling from the master sample at a constant rate.

An alternative, which is more appropriate when the units used as master sample USUs vary appreciably in size, is to select the master sample USUs with varying probabilities and to vary the sampling rates within master sample USUs to make the resulting survey sample Sampling within the USUs is easier if the required self-weighting. sampling intervals are always integers, i.e., take in 1 in 3, 1 in 5, 1 in 10, etc. As was explained in section C,2,c, this goal can be met by using a master sample design which assigns measures of size to all sampling units equal to their expected number of ultimate clusters, rounded to the nearest integer. This is a good example of how the use of a somewhat more complex selection scheme for the master sample itself can simplify the selection of samples from master sample USUs. The technique is illustrated by the master sample designs for Australia, Botswana, and Jordan (see the respective case-studies).

A master sample can also be designed to simplify estimation of sampling errors for surveys based on it. The use of a master sample containing several replicates or divided into random subgroups makes possible the use of relatively simple variance estimators, provided two or more replicates or random subgroups are included in the sample for each survey. Another master-sample design feature that can simplify variance estimation is the independent selection of two or more PSUs from each stratum used in the master sample design. For the most part, however, this and other design features that simplify variance estimation cost something in terms of sampling efficiency because they restrict the use of deep stratification and systematic selection procedures. A detailed examination of these issues is beyond the scope (Chapter VIII of Technical Paper 40 (U.S. of this technical study. Bureau, 1977) provides a useful description of variance Census estimation procedures used in a multiround household survey.) The point to be made here is that a master-sample designer should consider the implications of proposed designs for variance estimation and should incorporate features that will simplify it, provided the cost in terms of sampling efficiency is acceptably low.

Last, but not least, a well-designed operating system for the storage, use and maintenance of the master sample can substantially simplify its use. The detailed selection procedures and the selection probabilities for all of the master sample units at every stage must be fully documented. A careful accounting should be kept of the units that have been included in samples and the surveys for which they have been used. Such an accounting will help to avoid excessive response burden and the problems associated with it. To the extent possible, the entire system should be computerized; however, one should avoid being put in the position where rapid selection of samples for unanticipated surveys is delayed by a shortage of computer programming personnel.

2. Case-study illustrations

The majority of the countries included in the case-studies use or plan to use their master samples for multiple surveys. Unfortunately, the documentation available for many of them lacks detail on the relationships among the samples to be selected from the master sample for the different surveys. Therefore, it will not be feasible to pick out two or three designs and compare their main features, as was done in section C for the use of master sampling principles in multiround surveys. However, there is one design feature for which the case-studies illustrate a wide range of alternatives, namely, the amount of sample overlap between surveys at the USU level. To what extent are the same households interviewed in different surveys?

Strategies go all the way from extensive overlap to deliberate avoidance of overlap. Following is a brief description of the strategy adopted by each country for which the documentation provided relevant information:

<u>Australia</u> - Overlap between USUs is deliberately avoided. Australia has a continuing Monthly Population Survey. Supplementary surveys on different topics are carried out sequentially, with some gaps, during the 5-year intercensal period. Samples for all surveys are selected from a master sample of census EAs (called CDs in Australia); however, the samples for the supplementary surveys are selected from different blocks (area segments) within the master sample EAs. For some of the supplementary surveys all of the master sample EAs are used, for others only a subsample.

Ethiopia - The amount of overlap was greater than for any other country included in the case-studies. Current agricultural surveys and crop production surveys are conducted annually. Sample holdings for the crop production surveys are a subset of the sample holdings for the current agricultural surveys in each PSU. Other topics, such as demographic characteristics, labour force, income, expenditures, nutrition and health are covered by one-time or periodic surveys. For most of these surveys, the samples of agricultural households were essentially the same as those used in the current agricultural surveys.

<u>India</u> - India uses a new master sample of census blocks and villages for each annual round of its National Sample Survey. Initially, information was collected from the same sample households on several broad topics (deliberate overlap). However, this approach was dropped, largely because of respondent fatigue. In some of the subsequent rounds, separate samples of households were selected from the master sample blocks and villages for each survey conducted during the round. Of late, however, the pattern has undergone further change: a major subject or group of related subjects has been selected for each round (e.g. employment and consumer expenditure in the 32nd and 38th rounds, population births and deaths in the 39th round), and all the information collected on a common sample of households..

Sri Lanka - The amount of overlap is not controlled. It is planned to use the same set of PSUs for two consecutive annual surveys, each covering a different set of topics. For the first survey, a systematic sample of housing units would be selected from the listing for each of the master sample PSUs. For the second survey, the housing unit listings would be updated and a new sample Various options for selecting the new sample from the selected. updated listings were under consideration: a simple random sample, a systematic sample with a random start determined without reference to the one used in the first year, or a systematic sample with a random start chosen as far away as possible from the one used in the first year. None of these options would result in a large amount of housing unit overlap, but overlap would be minimized by the third option.

Thailand - A different type of overlap is illustrated by Thailand's use of its master sample. A migration survey, whose target population consists of households with one or more in-migrants, is conducted annually in the Bangkok Metropolis. Households with in-migrants are identified by the use of suitable screening questions during the listing of households in sample PSUs for the annual Labour Force Survey. All such households are included in the migration survey. Sample households for the Labour Force Survey are chosen at random without reference to their migration status, so there is a moderate amount of overlap in the samples of households selected for the two surveys, affecting only households with in-migrants.

The arguments favouring overlap of sample USUs for different surveys are weaker than they are for overlap in successive rounds of a single survey: the reliability of estimates of change is no longer a consideration. In theory, if two surveys are conducted at about the same time for the same sample of elementary units, the records for individuals or households from the two surveys can be linked to form a richer database for analysis. In practice, such linkages are not made very often, especially in developing countries. A better strategy, if one wants a database for multivariate analyses involving separate topics, is to collect as much of the data as possible in a single survey, thus avoiding the very real technical difficulties of linking records from different surveys.

On the other hand, the main arguments against overlap, i.e., the need to avoid excessive burden on survey respondents and the dangers that conditioning effects may cause the quality of response to deteriorate, apply with almost the same force to multiround surveys and multiple surveys.

One would probably have to conclude that the USU overlap between surveys in Ethiopia was excessive. Most of the agricultural households were interviewed on as many as 40 separate occasions during a period of two years! The other countries for which the sample overlap between surveys could be determined had none or only a moderate amount. Some overlap can certainly be justified if, as in Sri Lanka, it simplifies procedures for sampling from master sample USUs. In Thailand, the overlap made it possible to economize by using the household listings for a national survey of the general population (the Labour Force Survey) to obtain a sample for a separate survey aimed at a restricted population: in-migrants to the Bangkok Metropolis.

It is, of course, quite important that a policy on USU overlap be established when the master sample is being designed, sirre the amount of overlap expected will be an important determinant of the size and structure of the master sample.

3. Some important design issues

Many of the important aspects of the design of master samples have been discussed earlier in this chapter: the number of stages of sampling, the types of units to be preferred at each stage, the use of self-weighting samples, the exhaustion of sampling units, the duration of use of the master sample and the amount and kind of overlap in different samples selected from the master sample. However, there are three questions that have already been alluded to but are important enough to discuss here in order to assure that they are dealt with adequately. These questions relate to sample size, replication and updating procedures. a. <u>How large should the master sample be</u>? A master sample should be large enough to provide samples for all or most of the surveys that are part of an IHSP. In addition, it should include a reserve that can be used for samples to meet unanticipated requirements.

The following four-step process is suggested as a basis for arriving at a rough estimate of the size of master sample needed. It is assumed that a decision has already been reached on the number of stages of selection and the kinds of units to be used at each stage.

<u>Step 1.</u> Identify the smallest geographic areas for which separate estimates will be required (publication areas) in any of the planned surveys. For many countries, estimates will be desired in some surveys for regions, states, or provinces. Separate estimates may be needed for the urban and rural sectors, either at the national level or within each region, state or province.

<u>Step 2</u>. Decide on the number of master sample units that will be needed at each stage to produce estimates of the desired precision, in a single survey or survey round, for each separate publication area. Procedures for determination of sample size for a single survey are discussed in the standard sampling texts, e.g., Kish (1965). In this step, the procedure and rates for sampling within master sample USUs will also have to be considered.

<u>Step 3.</u> Decide on the system of overlap between different surveys and survey rounds. If the master sample USUs are to be selected in a single stage, will the requirements for different surveys be met by using different master sample USUs, by selecting new units in the same USUs, or by some combination of the two methods? If the master sample units are to be selected in two stages, it may be possible to meet all requirements from the same set of master sample PSUs, using different or partly overlapping sets of master sample SSUs in these PSUs.

<u>Step 4.</u> Use the results of the first three steps to arrive at an estimate of overall size requirements for the master sample. Expand the estimate from Step 2 for a single survey or round to allow for sample rotation in a multiround survey (if applicable), additional surveys included in the IHSP, and a reserve adequate to provide samples for surveys not initially included in the plan.

For further guidance, Exhibit 5.7 shows the single survey sample size requirements for the smallest publication areas in five countries. Each of the countries shown in the exhibit requires survey estimates for subnational areas. One-stage master-sample designs are used by Ethiopia, India and Nigeria, while Morocco and Saudi Arabia use two-stage designs. Thus, the master sample USUs are PSUs in the first three countries and SSUs in the last two.

Country and sector (1f	Smallest		Average number of master sample units for smallest publication area			
applicable)	Number	Туре	Number	Туре	Number	Туре
Ethiopia	12	Regions	4 <u>21</u> /	Farmers associations	NA	· · · · · ·
India	31	States	430 <u>2</u> /	Villages and urban blocks	NA	
Мотоссо						
Urban	7	Regions	56	EA clusters	195	EAs
Rural	7	Regions	20	Commune groups	50	EAs
Nigeria						
Urban	19	States	40	EAs	NA	
Rural	19	States	30	EAs		
S. Arabia	5	Regions	4	Emirates	60	Area segments

Exhibit 5.7 Number of master sample units per survey or survey round for smallest publication area: selected countries

NA - Not applicable, i.e., master sample has only one stage.

- EA (Census) enumeration area. $\frac{1}{2}$ With a minimum of 34 in any region.
- $\frac{1}{2}$ A parallel sample is available for states wishing to obtain substate estimates. The number of PSUs varies from 12 to 1500 depending on the size of the state.

The average number of master sample PSUs per publication area for Ethiopia, Morocco and Nigeria varies within a fairly narrow range, 20 to 56. The much higher range for India probably reflects the desire by some of the states to produce separate estimates for substate areas or domains. The Indian National Sample Survey is а cooperative federal-state enterprise and emphasis is on meeting the needs of the The design for Saudi Arabia goes to the other individual states. extreme, and it is questionable whether the estimates at the regional level would be sufficiently reliable to be of much value. Indeed, the national sample of 21 PSUs (emirates) and about 300 segments in these 21 PSUs is perhaps marginally sufficient to produce useful estimates. No matter how large a sample is selected within the 300 segments for a particular survey, the sampling errors are likely to be dominated by the between-PSU and the between-segments within-PSU components of variance. To put it another way, the design effect from clustering the sample is likely to be quite large.

Usually the actual selection of the master sample units will be a relatively inexpensive undertaking, since the work will be carried out in the central office and will require only a moderate amount of professional, clerical and computer time. The larger costs are incurred in connection with the preparation of maps and SSFs for the sample units. For this reason, the size of the master sample should be large enough to meet all foreseeable meeds. If there is any possibility that separate samples for states or provinces may be needed, it costs little to provide for them, even though they will not be used at an early stage. If the master sample is being selected from a recent census frame, census materials, such as EA maps and listings, for master sample EAs should be obtained as soon as the selection has been made. However, subsequent steps needed to prepare SSFs for master sample USUs should be taken only for those units which are about to be used in a scheduled survey.

In summary, the recommended strategy for determining the size of a master sample is to make it large enough so that there is virtually no chance of being caught short during the planned period of use. This strategy adds very little to the actual selection costs. The more costly activities associated with actual use of master sample units can be deferred until the units are about to be used.

b. Is the use of replication desirable? The experts who reviewed the initial outline for this technical study had varying views about whether or to what extent the use of replication in a master sample is desirable. Furthermore, there is considerable variation in the sampling literature with respect to precise meanings assigned to the term replication and closely related terms: interpenetrating subsamples, random subgroups and pseudo-replication. It would be unreasonable to expect that this technical study can resolve these complex issues once and for all. However, it may be possible to clarify the terminology and the issues, and to provide some general guidelines. Following, approximately, the usage in a previous NHSCP technical study (United Nations, 1982a), the term interpenetrating subsample will be used to describe one of a set of subsamples, all of which are part of a larger sample and each of which constitutes, by itself, a probability sample of the target population. The subsamples may or may not be selected independently of each other. If they are selected independently at all stages they will be referred to as replicates; if not, they will be called random subgroups (in the literature they are variously referred to as random subgroups, rotation groups, panels, psuedo-replicates or subsamples, depending on the context).

A simple example will serve to illustrate the difference between replicates and random subgroups. Consider a one-stage master sample that will consist of census EAs, to be selected with equal probability, without stratification, using a systematic selection procedure. A master sample of 2,000 EAs is desired and (for reasons to be discussed shortly) it is also planned that the master sample will consist of 20 interpenetrating subsamples, each consisting of 100 EAs.

In this illustration, the interpenetrating subsamples could be either replicates or random subgroups. To select 20 replicates, one would select 20 random starting points and apply the sampling interval N/100, where N is the total number of EAs, to each of the random starts. If a particular random start were chosen more than once, the replicates corresponding to that random start would be identical.

To select 20 random subgroups, one would start by selecting a single systematic sample of EAs with a random starting point and a sampling interval N/2,000. The resulting sample EAs would be numbered consecutively from 1 to 20 in each successive set of 20 sample EAs. The random subgroups would consist of all EAs assigned the same number, i.e., all 1s, all 2s, etc.

Why should a master sample be designed to consist of interpenetrating subsamples, whether replicates or random subgroups? There are several reasons:

- o The subsamples provide flexibility with respect to sample size. Depending on data requirements, the sample for a particular survey or survey round can consist of one, two, or more subsamples.
- o The subsamples can be used for sample replacement in successive rounds of multiround surveys. For example, the first-round sample might consist of four subsamples, with one subsample to be replaced by a new one after each round.
- In a survey whose sample consists of two or more subsamples, relatively simple variance estimators based on subsample totals can be used.

o The subsamples can be used in various ways to measure and control non-sampling errors (see United Nations, 1982a, pp 179-180 and 239-243).

The illustration given above divided the master sample USUs (census EAs) into interpenetrating subsamples. However, it is also possible to create interpenetrating subsamples in the process of subsampling within a fixed set of master sample USUs. Using the assumptions of the earlier example, one could start with a single sample of 2,000 EAs, prepare housing unit listings for each EA, and distribute the housing units in each EA among 20 separate clusters. In this case each random subgroup would consist of one-twentieth of the housing units in every sample EA.

It would <u>not</u> be possible in this last example to create full replicates, since the sample of PSUs is assumed to be fixed. However, it would be possible to select two or more conditionally independent subsamples, each consisting of one-twentieth of the housing units in every one of the 2,000 sample EAs. These interpenetrating subsamples might be referred to as partial or conditional replicates.

From the case-studies and from general consideration of how master samples are likely to be used, it is clear that interpenetrating subsamples of some kind are almost certain to be used in an IHSP based on a master sample. Which type of interpenetrating subsamples, replicates or random subgroups, should be preferred?

Full replicates have one clear advantage: they minimize the number of assumptions or adjustments needed to estimate sampling errors. Normally, with a sample consisting of two or more full replicates, unbiased estimates of sampling errors may be obtained by using each replicate to estimate a total or other statistic and then calculating the variance between these estimates. This explains why a sample design with two PSUs per stratum is often used: it provides two full replicates, each consisting of one PSU from every stratum.

But there are also disadvantages to using full replicates. For one thing, they restrict the use of sample selection techniques that may lead to more efficient designs, e.g., deep stratification (one PSU per stratum), controlled selection and fully systematic sampling, Suppose, in the general framework of the earlier example, a set of four interpenetrating subsamples, each consisting of 5 EAs, were needed. If random subgroups were used, the sample EAs in each subgroup and in the full sample would be equally spaced in the population, i.e.:



However, if replicates were used, each replicate would require a separate random start, and a possible result would be:

Thus, the EAs in each of the replicates would be equally spaced, but for the full sample they would not be. As a rule, the second configuration of the sample could be expected to result in some loss of efficiency.

A more important consideration, however, is the additional cost that might be introduced by using full replicates instead of random subgroups. As pointed out earlier in this chapter, the greatest economies of scale in connection with the use of master samples are likely to come from making the greatest possible use of SSFs developed for master sample USUs. Therefore, the introduction of new replicates for purposes of sample rotation in a multiround survey before exhausting at least a substantial proportion of the SSFs developed for the replicates used in earlier rounds could be inefficient, especially if the estimation of change is an important survey objective.

There are, of course, ways in which replicates can be used efficiently. For example, one could use a set of full replicates in one survey and then use the same PSUs (say census EAs) in a second survey, but select new samples of segments or housing units from the SSFs already created for the sample PSUs. Each of the two surveys would have a full replication design, while sharing between them the cost of SSF development.

In summary, interpenetrating subsamples play an important role in the efficient use of master samples. The choice between replicates and random subgroups depends on various technical considerations. The advantages of using full replicates for variance estimation should not be purchased at too high a price, i.e., if a significant increase in the cost of preparing SSFs is necessary.

c. <u>Updating</u>. Updating of samples and sampling materials for an IHSP occurs at three different levels. In most countries the master sampling frame (MSF) is fully updated after each population census. In some countries less extensive changes are made in the MSF between censuses to reflect new administrative divisions and subdivisions and population growth that is unevenly distributed.

At the next level, the period for which a master sample is used varies from as little as one year to the full intercensal period. Some countries adopt a sample replacement strategy, i.e., an entirely new master sample is selected after a relatively short period. Others adopt a sample revision strategy: the same basic master sample is retained over the full intercensal period, but new sample units are added at one or more levels to reflect uneven patterns of growth. Finally, it may sometimes be necessary to update the secondary sampling frames (SSFs) developed for master sample USUs. In particular, SSFs consisting of list units -- housing units or households -- age quickly and usually require updating after not more than one or two years.

All of these updating strategies and the procedures associated with them have been discussed earlier in this technical study. It may be useful now to summarize current country practices, as revealed by the case-studies. There were some countries for which the available documentation did not cover updating procedures: this does not necessarily mean that no provisions for updating have been made.

<u>Australia</u> - Administrative data on building approvals (permits) are used to update the MSF twice a year. For the master sample, a sample revision strategy is followed: additional master sample USUs are selected in strata with large growth. SSFs are created in two stages: first the master sample USUs are divided into area segments called blocks, then the sample blocks are listed and the dwellings allocated to a specified number of clusters, usually from 4 to 8. If a master sample USU is found to have large growth relative to the stratum from which it was selected, the division into blocks is redone and new sample blocks are selected. Listings within blocks do not require updating, since the period of interviewing within a block, allowing for monthly rotation of one-eighth of the sample clusters, does not normally exceed 15 months.

Botswana - In practice, the proposed master sample was used for only one survey. For many of the master sample USUs, lists of dwellings from a 1981 population census were apparently used without updating as SSFs for a survey conducted in 1983.

Ethiopia - The master sample, consisting of 500 farmers' associations, was used without updating over a four-year period. Household listings were prepared for the master sample USUs at the start of the four-year period and again after two years. The documentation does not say whether the second set of listings were done independently or were updates of the first listings.

<u>India</u> - The MSF, which consists of a listing of census blocks and villages, is said to be updated periodically to reflect boundary changes. For each annual round of the National Sample Survey, a new sample of blocks and villages is selected. Household listings are prepared for use as SSFs just prior to the start of the round and are used only for that round. For this reason, no updating of the household listings for the master sample USUs is required.

<u>Jordan</u> - In the proposal for the master sample there are no provisions for updating it or the MSF. The SSFs for the master sample USUs (census blocks or block groups) were to be census listings of housing units, updated prior to use. <u>Morocco</u> - Provisions for updating are not covered by the master sample proposal.

<u>Nigeria</u> - The MSF consisted of a list of EAs from a 1973 census, without measures of size. A master sample of EAs was selected for use in an IHSP starting in 1981. The documentation (a 1985 report) mentions work underway to update the census EA frame in preparation for an upcoming census and recommends selection of a new master sample when the updated frame is available. With the present master sample of EAs, one-fifth of the sample EAs are replaced each year. At the start of each year a new listing of households is prepared for every EA, whether or not in sample the previous year.

<u>Saudi Arabia</u> - The documentation does not discuss updating procedures at any level.

<u>Sri Lanka</u> - The MSF, a listing of EAs from the 1981 population census, was updated late in 1984 to account for changes in administrative divisions and subdivisions and uneven patterns of growth resulting primarily from economic development programmes. A master sample of EAs was selected from the updated MSF for use in a survey starting in April 1985. The same sample of EAs is to be used again for the survey scheduled for 1986/87. The SSFs for the master sample EAs are housing unit listings. They are to be updated for each EA prior to the scheduled inclusion of that EA in the following year's survey.

Thailand - The MSF for the urban sector is a listing of blocks (defined areas within EAs) based on the 1980 population census. It has not been updated. The MSF for the rural sector is a list of villages, which is updated annually, based in part on information from the government agency responsible for creating new villages and in part on an annual survey of villages conducted by the statistical office. A new master sample of blocks and villages is selected each year and a housing unit listing is prepared for each master sample USU shortly before its first scheduled use.

<u>United States</u> - The documentation does not tell what procedures, if any, were used to update the MSF and the master sample. Updating requirements were likely to have been minimal, for two reasons. First, the master sample USUs and the SSF units were both area segments with well-defined boundaries. Second, the elementary units for most of the surveys that used this master sample were either farms or tracts of farmland. Since both the number of farms and the total area of farmland have been decreasing in the United States over the past several decades, unusual growth of sampling units was not a likely occurrence.

This recital of current practice should serve to reinforce some points that have been discussed earlier. First, it is clear that strategies for master sample updating vary along a broad spectrum, from the sample replacement strategy with each master sample used for a one-year period (e.g., India, Thailand) to the sample revision strategy, with the same master sample being used for as long as five years (e.g., Australia). Either strategy can be used successfully. The primary trade-off to be evaluated in making a choice is between the design simplicity that is possible when the replacement strategy is used and the somewhat greater economies of scale that can be realized by using the revision strategy.

Another observation is that most countries that use list SSFs for the master sample USUs prepare the listing for each USU just prior to the first appearance of that USU in a survey. In most but not all countries the list SSFs are not used for more than one year without updating.

Finally, the failure of the documentation for some countries to include provisions for updating may indicate that some IHSP designers are giving insufficient thought to this critical design element. Provisions for updating should be a part of every IHSP design from the beginning: such advance planning will avoid troublesome complications that are likely to be encountered if updating procedures are developed only at some later time when it becomes obvious that they are needed.
E. <u>Special topics relating</u> to master samples

1. Using a master sample in combination with other samples

The main objective of a master sample should be to provide samples for household surveys that are part of an IHSP and that have reasonably compatible design requirements with respect to publication areas and distribution of the target population within those areas. Some reserve capacity for unanticipated household survey needs is desirable.

However, to go beyond this point and try to design a master sample to provide complete samples for a wide variety of non-household surveys and local area surveys may lead to diminishing returns. The design will become increasingly complex and it may be necessary to make compromises that render the overall master sample design less efficient for the basic set of household surveys.

An alternative that may work better is to use the master sample in combination with samples drawn from other sources for the surveys for which the master sample alone cannot provide an efficient sample. There are two possible procedures: a single-frame and a multiple-frame procedure.

The single-frame procedure relies entirely on the MSF from which the master sample was selected. The general approach would be to use existing master sample units to the extent they are available and will not be needed for other scheduled surveys, and to meet the remaining requirements with additional units of the same kind selected directly from the MSF. Two simple examples will show how this method might be used.

For the first example, suppose that a rather large sample of census EAs is needed for a special household survey in a single metropolitan area. The master sample, which is a one-stage national systematic sample of census EAs, can provide about half of the EAs needed for the special survey and unused area segments are available in each of these EAs. The remaining half of the census EAs needed would be selected directly from the MSF. For the new sample EAs it would be necessary to prepare SSFs. This could be done either by preparing household listings or by segmentation using existing maps or EA sketches, depending on which method was judged to be more efficient for the purpose of the special survey.

For the second example, suppose that a sample is needed for a survey of blind persons. The existing master sample in this case happens to be a two-stage national sample of census EAs, with districts (small administrative subdivisions) serving as PSUs. Regular interviewers will be available in the master sample PSUs to do screening to locate blind persons and to interview those found. Preliminary analysis suggests that an ultimate cluster larger than a single census EA would be optimum for this survey because the average EA would be expected to contain only 1 or 2 blind persons.

In this example, one might decide to use all of the master sample PSUs and, using the MSF, divide each one into clusters of about 5 adjacent census EAs. One or more of these clusters would be selected for the special survey. All households in the sample clusters would be listed and screened for the presence of blind persons. The main advantages in this case would accrue, first, from the use of a suitable sample of PSUs that had already been selected and, second, from the use of experienced field-workers stationed in the master sample PSUs.

The multiple frame approach relies in part on the master sample and in part on frames other than the MSF. The case-study for Ethiopia provides one example. The target population for the current agricultural survey consists of all agricultural holdings, including state farms, cooperatives and private holdings. Data for all of the state farms, without regard to location, are obtained from the Ministry Data for cooperatives and private holdings are of State Farms. collected only in the master sample USUs. All cooperatives in the master sample USUs are included in the sample, but only a subsample of the private holdings. The basic frames in this case are the MSF and the ministry's list of state farms. For the master sample USUs, two sets of SSFs are used: lists of cooperative farms (probably obtained from local authorities) and lists of private holders identified in the course of household listing operations.

As mentioned previously (see subsection B,2 of this chapter), the Ethiopian example can be generalized to cover various types of economic surveys. In most developing countries a large proportion of the economic establishments, especially in the agricultural, retail trade and service sectors, are directly associated with private households. These establishments can be surveyed with reasonable efficiency within the framework of a master sample developed for household surveys. However, there are usually some large establishments not directly associated with private households. They may be small in number but account for a large proportion of total employment and output. They are also likely to be unevenly distributed with respect to the general population.

Separate list frames are needed for these large units. The lists can be developed through records of government agencies, from prior economic censuses and surveys, or by field canvasses. Using these lists, a sample design frequently used divides the target population of economic establishments into four size categories and uses a different sample strategy for each category:

	Size category	Sampling procedure
1.	Small (associated with private households)	Sample of households in master sample USUs
2.	Intermediate (separate list for master sample USUs	Take all in master sample USUs
3.	Large (separate national list)	Sample without regard to location
4.	Very large (separate national list)	Take all without regard to location

There are many other ways, in addition to those illustrated above, in which either a single- or multiple-frame sample design can be used to permit some part of the sample requirements for a survey to be met from a master sample. For example, suppose a national sample of physicians is needed. The national medical association has a list with current addresses for an estimated 75 percent of all active physicians. Depending on sample size requirements, the sample might include all physicians from this list in the master sample USUs, or a subsample of them. A sample of physicians not on the list could also be obtained in the master sample USUs, either by a special screening operation or by screening in ongoing surveys.

The lesson to be taken from these illustrations is that whenever needs arise for a new survey, the first question that should be asked is whether the MSF, the master sample and the SSFs currently available for master sample USUs can provide part of the sample for that survey. Frequently the answer will be yes, and the resulting savings can be substantial. Multiple-frame sampling is a useful tool that makes many of these efficient designs possible.

2. Can master samples for household surveys be used for agricultural surveys?

The answer to the question should be obvious by now: it is yes, up to a certain point. The previous subsection described how Ethiopia uses its household master sample to represent all agricultural holdings other than state farms. Private holders are identified as part of the listing operations in master sample USUs and separate listings of cooperatives are developed for the same USUs. In Ethiopia the agricultural surveys are, in fact, the core element of the IHSP. This is also true in several other African countries: Kenya, Lesotho, Malawi, Mali, Zambia and Zimbabwe (United Nations Statistical Office, 1983). As described in the case-study for Nigeria, data on farm characteristics, livestock, crop areas and production are collected annually as part of that country's IHSP.

-174-

In India crop-cutting surveys were carried out in the master sample USUs selected for each annual round in the early years of the National Sample Survey. Prior to the period covered by the case-study for Sri Lanka, that country collected data on private agricultural holdings in a survey of household economic activities. However, in Sri Lanka, annual data on area and production of rice are collected in a survey whose design is entirely independent of the household survey programme.

The extent to which agricultural statistics and household survey activities can be integrated in a country depends on several factors: the structure and geographic distribution of agricultural activities, the kinds of agricultural data to be collected and the existing institutional arrangements for agricultural statistics. In most of the developed countries, only a small proportion of the population are engaged in agriculture and many holdings are not directly associated with households. In these circumstances, a master sample designed for multi-subject household surveys would generally not be useful in connection with agricultural surveys. Thus, the case-studies indicate that the United States developed a master sample specifically for use in agricultural surveys and Australia does not use its household survey master sample for agricultural surveys.

In many developing countries, however, the agricultural sector is composed largely of small family-operated holdings and a large proportion of the households, outside of large cities, include one or more holders. In these circumstances, partial integration of household and agricultural surveys is well worth considering.

Much depends on the kinds of data to be collected. For some kinds of data it is essential that the holding be used as the primary unit of observation. This is generally the case for topics such as size and type of holding, livestock inventories and production, agricultural equipment, agricultural practices and labour inputs. For these topics, the information must be obtained directly from the holder and the household survey approach is a natural one to use. For data on crop areas and production, it is generally agreed that objective measurement techniques, including actual harvesting and weighing of crops in sample plots, are needed to obtain data of acceptable accuracy. If reasonably accurate land records are available, efficient samples of fields and plots can be selected directly from these records, by-passing the household as a sampling unit (of course arrangements must be made with the holders to make the measurements on their land). However, where good land records are not available, as in many African countries, a household survey may provide the best framework for the selection of a sample of fields from holdings associated with the sample households.

There are often a cumber of operational, technical and institutional difficulties to be overcome in integrating household and agricultural surveys. In particular, in most countries systems of agricultural

statistics and multi-subject household surveys have evolved as distinct entities, often in different ministries or departments, so that there may be strong resistance to integration. Nevertheless, opportunities for integration should be fully explored. As pointed out by the United Nations Statistical Office, "...it is neither possible, nor indeed necessary ... to maintain two parallel systems of statistical data collection dealing practically with the same set of rural households" (1983, p. 1370).

3. Treatment of special population groups

Special population groups are mainly of two kinds: persons living in institutional settings and nomadic or tribal groups. For various reasons, these groups are difficult to cover with the sampling and data collection techniques normally used in household surveys.

The case-studies do not provide any illustrations of how master-sampling principles might be used to cover nomadic and tribal groups in household surveys. Ethiopia and Jordan explicitly exclude nomadic groups from their survey target populations. Morocco excludes rural population in sparsely populated desert areas, roughly 10 percent of the total rural population. The documentation for Saudi Arabia makes no reference to the treatment of nomadic groups. Thailand explicitly excludes hill tribes from the regular household surveys: these groups have been studied from time to time in special surveys. India, however, covers tribal areas in the same way as other areas.

There are undoubtedly good reasons for excluding some nomadic and tribal groups from national household surveys. No population group, however, should be entirely excluded from a country's statistical programmes. Their inclusion in population censuses is desirable, if feasible, as well as occasional special surveys designed specifically for such groups. With respect to sampling methods, in countries where nomadic and tribal groups have distinct identities it would be useful to create and maintain a list of such groups that could serve as a master sampling frame. The frame units would be the smallest groups that are separately identifiable, and the record for each unit would include a measure of size and information that would help to locate the group at any time.

In similar fashion, most of the case-studies for developing countries do not describe procedures for sample coverage of persons living in institutional settings. In most cases where the target population was explicitly defined, the institutional population was excluded.

Australia, however, includes persons living in institutions in its target population for household surveys. For sampling purposes a distinction is made between private dwelling units and special dwelling units (SDs). Included in the latter category are hotels, hospitals, prisons, construction camps, etc. Separate sampling frames for SDs are established and they are updated twice a year. In densely populated areas (where a one-stage master sample of census EAs is selected to cover the private dwellings) a universe list of SDs is maintained and samples of SDs for monthly survey rounds and supplemental surveys are selected directly from that list, with no master sample as an intermediate stage. However, in the sparsely populated areas, the SD lists are maintained only for the master sample PSUs, which are groups of adjacent census EAs.

In general, the special population groups are apparently not being covered in most household surveys in developing countries. As IHSP designs become more sophisticated, it may be possible and desirable to include some of these groups, especially the institutional population, in some surveys. At such time, the broad guidelines, set forth in this study, for the development and use of MSFs and master samples should be of some value. Occasionally, as illustrated by the Australian case-study, a master sample developed for coverage of private households may also be used as an intermediate stage in sampling the institutional population.

4. Quality assurance

One of the advantages of using a master sample, pointed out earlier in this chapter, is that some of the savings from re-using master sample units and the SSFs prepared for those units can be invested in improving the initial quality of the design and the sampling materials. Such an investment in quality should, in fact, be more than just a possibility; it should be considered a requirement. Once a master sample has been selected, the fact that it will be used for several surveys or survey rounds should lead to appreciation of the importance of preserving and updating all materials associated with the master sample units and recording accurately the details of all uses of the master sample.

Appropriate techniques for assessing and controlling non-sampling errors in household surveys have been presented in detail in an earlier NHSCP technical study (United Nations, 1982a). Many of these techniques are applicable to the design and selection of a master sample. A few that should be especially useful will be mentioned here.

First, all sample selection operations at every stage of selection should be checked, both for the master sample itself and for subsamples selected from it. Checking procedures may include:

- Independent repetition of the sample selection procedure, using identical selection probabilities and random selection points. The resulting samples should be identical to those selected initially.
- Checking actual sample sizes against expected values that have been calculated in advance by applying selection probabilities or intervals to frame counts.

 Using variables available for master sample units (and not used in the determination of selection probabilities) to estimate corresponding population totals. For example, if a master sample of EAs has been selected with equal probability (at least within strata), census counts of persons for the sample EAs can be used to estimate actual census population counts. If the master sample (or subsample) consists of two or more replicates or random subgroups, the estimates can be made separately for each one.

With one or more of these checks, it should be possible to spot any gross errors in the sample selection operations.

Second, a standard system of unique numeric identifiers should be developed for the master sample units at each level and for all units in samples selected from the master sample. Desirable features of a system of identifiers of sampling frame units were discussed in subsection D,3 of Chapter IV. Similar principles apply to identifiers for sample units. An additional consideration is that the identifiers should be designed to facilitate the calculation of sample estimates and their sampling errors. It should be possible to sort on one or two digits of the identifiers to group data relating to the same publication area. If sample revision strategy is to be followed, the system the of identifiers for the master sample units should allow for units that may be added as the result of updating.

Third, as in all survey operations, good documentation is essential to the effective use of master samples. The documentation should include, in particular, detailed descriptions of the sample selection procedures for the master sample and all samples selected from it. All sampling worksheets and computer printouts used in the selection process should be preserved. All uses of the master sample should be recorded so that it will be possible to determine, for each master sample unit, how often it has been used and in which surveys.

Much more depends on the quality of a master sample than on the quality of a sample that is to be used for only one survey. The use of a master sample represents an opportunity to invest more resources in good initial quality. It also entails a responsibility to realize the benefits of investment by checking and documenting fully all applications of the master sample.

-179-

CHAPTER VI

SUMMARY AND RECOMMENDATIONS

A. Summary

The history of population censuses goes back to ancient times. Sample surveys are a more recent development. The mathematical foundations of sampling theory were developed in continental Europe and Great Britain in the 19th and early 20th centuries. However, extensive practical applications of sampling theory in government surveys occurred only after the publication in 1934 of an historic article by Jerzy Neyman in which he developed three fundamental concepts: optimum allocation in stratified sampling, confidence intervals for sample estimates, and the importance of using random as opposed to purposive selection methods.

Following the publication of Neyman's article, methods of sampling from finite populations were rapidly developed, refined and applied by statisticians such as Cochran, Deming, Hansen, Hurwitz, Mahalanobis, Stephan and Yates. Most of the early applications in household surveys were for <u>ad hoc</u> surveys covering topics such as unemployment and consumer expenditures.

Before long, however, two things became evident: first, that continuing or periodic household surveys could be used to develop, at an affordable cost, reliable time-series data on topics such as labour force participation and unemployment and, second, that economies of scale and improvements in quality could be achieved by making use of the same personnel, facilities and sampling materials for more than one survey. In the United States, the first practical application of probability sampling for a continuing survey was in 1940 with the start of the Sample Survey of Unemployment, later to become the Current Population Survey (Duncan and Shelton, 1978). Shortly thereafter, as already described in subsection A,1 of Chapter V, the concept of a master sample which could provide samples for several different surveys was pioneered by the United States Department of Agriculture. In 1950, under the guidance of Mahalanobis, the National Sample Survey of India came into being as a permanent survey mechanism (Rao and Sastry, 1975).

In spite of these early examples, the benefits of an integrated programme of household surveys are not widely recognized. Most texts and manuals on sampling and survey methods still use one-time or <u>ad hoc</u> surveys as the primary context for the description of suitable survey designs and procedures. Prior to the launching of the National Household Survey Capability Programme (NHSCP), the tendency in many developing countries was to undertake household surveys on an <u>ad hoc</u> basis. Those countries that did establish continuing or periodic surveys did not always use designs that enabled them to realize the full benefits of integration. A basic requirement of the NHSCP has been that each participating country should develop an IHSP. There is no standard design for an IHSP: each country is expected to develop a design that is suited to its own data requirements and resources. There are certain principles, however, that can help each country to realize fully the potential benefits of an IHSP.

As described in section A of Chapter II, integration refers to linkages between surveys or survey rounds. These linkages relate to the standardization of survey content, the sharing of survey personnel and facilities, and the use of common samples and sampling frames. The main focus of this technical study has been on the last of these three aspects of integration, i.e., the development of sample designs for IHSPs.

The typical procedure for obtaining samples for individual surveys or survey rounds in an IHSP requires six steps:

- 1. Take a population census.
- 2. Use census materials to create a master sampling frame (MSF).

SAMPLING PROCEDURES	3.	Select a master sample from the MSF.
FOR AN IHSP	4.	Select a subsample of units from the master sample.
	5.	Create secondary sampling frames (SSFs) for the selected master sample units.

6. Select samples for individual surveys from the SSFs.

Steps 1 and 2 would be modified, of course, if no suitable materials from a recent census were available. Steps 5 and 6 could be omitted if the master sample units were small "take-all" area segments. However, the pattern shown is the one followed in most countries.

Specifications for the frames and samples used in this process are guided by the overall design for a country's IHSP. Chapter III describes three major classes of IHSP designs and the factors that should be considered to make an appropriate choice among these designs.

Three of the six steps in the sampling process relate directly to the construction of sampling frames, which is the topic of Chapter IV. Planning for the creation of an MSF (Step 2) should start prior to the population census (Step 1), to ensure that the census will provide the outputs needed to create the MSF. Section D of Chapter IV provides a detailed account of the process of creating and maintaining an MSF. Step 5 in the sampling process is the creation of SSFs in master sample units that have been selected for use in particular surveys or survey rounds. Procedures for creating area and list SSFs are reviewed in Section E of Chapter IV.

The other three steps in the sampling process -- steps 3, 4 and 6 -involve the selection of samples. In step 3 a master sample is selected from the MSF. In step 4, subsamples for use in one or more surveys or survey rounds are selected directly from the master sample and finally, in step 6, samples (usually of housing units, households or small area segments) are selected from the SSFs prepared for a sample of the master sample units. These three steps are discussed in Chapter V. The choice of design for a master sample depends largely on its intended uses. Section C of Chapter V covers the use of master sample concepts in the design of multiround surveys; Section D covers the use of master samples for different surveys.

The establishment of an integrated system of frames and samples is an essential part of the plan for an IHSP. The considerable costs of frame development and sample selection can be substantially lowered, on a per survey basis, by taking full advantage of population census outputs and by developing frames and samples that can be used for more than one survey or survey round. Some of the cost savings can be applied to improving the quality of the sampling frames.

The switch from the conduct of household surveys on an <u>ad hoc</u> basis to the operation of an IHSP requires a firm commitment to systematic long-range planning. The need for advance planning is illustrated by the six-step sampling process just described. Prior to the population census, the general structure of the MSF must be determined, with emphasis on the choice of basic frame units. The design of a master sample requires preliminary decisions on the number, timing and content of the surveys for which it will be used and on the general structure of the field staff for the survey operations. Early in the intercensal period, a decision must be made whether to use a sample replacement strategy, which calls for the selection of a new master sample after a relatively short period, or a sample revision strategy, which keeps the same master sample for a longer period, with periodic adjustments to reflect changes in the structure of the survey target populations.

The requirement for careful advance planning does not mean that the statistical office is irrevocably committed to a particular set of surveys and samples. One of the most important requirements in the design of master sampling frames and master samples is to build in some flexibility so that unexpected data needs can be met quickly, relying largely on facilities that have already been created.

The primary conclusion of this technical study is that master sampling frames and master samples are essential and valuable elements for integrated household survey programmes. The study has presented detailed recommendations for their use: the most important of these recommendations are summarized in the following section.

B. Recommendations

The main points that should be considered in the process of designing, selecting and using samples for an IHSP are summarized in the following check-list. Where appropriate, references are given to relevant sections of Chapters III, IV and V.

1. Choice of an overall IHSP design

- a. Before each population census, start planning for household surveys to be conducted during the next intercensal period (III,A,1 and 2).
- b. Make a preliminary choice of subjects and decide on the frequency of coverage for each (III,B,1 and exhibit 3.1).
 - (i) Decide which subjects can be grouped in the same survey or survey round (III, B, 2 and 3 and exhibits 3.2 and 3.3).
 - (ii) Examine possibilities for linking subjects from different surveys at the analysis stage (V,B,1).
- c. Make a realistic evaluation of the resources and staff available for field-work, data processing, and survey design and management (III,C,4,5 and 6).
- d. Decide which of the general classes of IHSP designs to adopt (III,D,3): a single multi-subject survey (III,D,1 and exhibit 3.5), two or more single-subject surveys (III,D,2), or a combination of these.
- e. Review the IHSP design annually and make changes as needed.
- 2. Design of a master sampling frame (MSF)
 - a. As part of planning for a population census, determine what outputs will be needed for use in constructing an MSF (IV,D, introduction and subsections 1 and 4).
 - b. Conduct a thorough review of potential inputs (primarily lists and maps) to the MSF (IV,D,2).
 - c. Decide on the basic frame units for the MSF (IV,A,2; IV,B; and IV,D,3). In making this choice, consider the following points:
 - (i) It may be desirable to have more than one type of frame unit, with a hierarchical structure, for example,

administrative districts and census enumeration areas (EAs) within districts (IV,D,1 and 3).

- (11) The units and structure of the MSF need not be the same for the entire country; sometimes urban and rural areas are treated differently (IV,D,1 and 3).
- (iii) Frame units with physically identifiable boundaries should be preferred, subject to the restriction that the units should not cross the boundaries of any publication areas to be used in surveys (IV,C,1).
- (iv) In most countries, census EAs are a good choice for use as one of the basic frame units (IV,D,3).
- d. Numerical identifiers for frame units have several important uses. The structure of the system of identifiers should be designed to facilitate all of these uses. The numbering system should allow for changes that may be necessary to reflect changes in frame unit boundaries (IV,D,3).
- e. Records for MSF units should provide for documentation of their use in master samples and other samples (IV,A,5).
- f. The MSF design should include a plan for updating. The nature of the updating requirements will depend to a considerable extent on whether a sample replacement or sample revision strategy is to be followed in using the MSF (IV,D,5).

3. Designing a master sample

- a. Determine the length of the period for which the master sample will be used. A sample replacement strategy implies a short period, probably not more than two years; a sample revision strategy implies a longer period, with periodic updates (IV,D,5; V,C,2,e and D,3,c).
- b. Identify surveys for which the master sample will be used (V,A,2 and D,3,a). Consider the possibility of using the master sample, alone or in combination with other frames, for surveys of economic establishments (V,E,1 and 2).
- c. Identify the smallest set of publication areas (areas defined for administrative or statistical purposes) for which separate estimates will be needed from any of the surveys for which the master sample will be used (V,D,3,a).
- d. Decide on the appropriate level and proportion of overlap between samples to be selected from the master sample for different surveys or survey rounds. The greater the overlap,

the greater the economies of scale that can be realized; however, the amount of overlap is constrained by possible adverse affects of excessive response burden and, in some instances, by the desire to accumulate aggregate data over survey rounds or subrounds (V,C,2,a).

- e. After considering points a to d above, determine how large a master sample is needed. Include reserve capacity for unanticipated survey needs (V,D,1).
- f. Design the master sample to avoid exhaustion of individual units. This is usually accomplished by combining units that are too small to meet expected subsampling requirements. For multiround surveys, if exhaustion of some units is expected, develop an unbiased procedure for replacement (V,C,2,d).
- g. If self-weighting designs are to be used for some surveys, design the master sample to facilitate the use of these designs (V,C,2,c).
- h. Use master sample design features that will facilitate the estimation of sampling errors for surveys whose samples are selected from the master sample (V,D,1). For this and other purposes, the use of a master sample consisting of interpenetrating subsamples may be considered; however, there are some limitations that should be taken into account (V,D,3,b).
- i. If the sample revision strategy has been adopted, develop a plan and schedule for updating the master sample (V,D,3,c).
- 4. Development and use of secondary sampling frames (SSFs) for master sample units
 - a. Where resources permit, use area units in preference to list units (IV,E,1 and 3 and exhibit 5.6).
 - b. If use of list units is necessary and the SSFs are to be used for surveys or survey rounds carried out at different times, use housing units in preference to households (IV,E,2 and 3).
 - c. SSFs with list units should not be used for extended periods without updating (IV,E,2).
 - d. Consider the possibility of placing serial number labels on structures or housing units (IV,E,2).
 - e. Establish clear rules of association (IV, B):
 - (1) Between master sample units and SSF units.
 - (11) Between SSF units and elementary units.

- f. Avoid long intervals between preparation of SSFs and their first use: prepare SSFs for master sample units only as needed (V,D,3,a).
- g. For maximum economy, use the SSFs prepared for master sample units in as many surveys or survey rounds as possible (V,B,1 and exhibit 5.2).

5. General considerations

- a. Develop full and accurate documentation of procedures and results at all stages of the frame development and sample selection process (III,A,5).
- b. Re-use of master sampling frames and master samples means that their quality assumes special importance. Quality assurance techniques should be made a part of every step in frame development and sample selection (see especially IV,D,2 and V,E,4).

-186-

ANNEX I

CASE-STUDIES

This annex consists of case-studies describing the design and use of master sampling frames and master samples for survey programmes in 11 countries. These case-studies are referred to frequently throughout the text, especially in Chapters III, IV and V. They were included so that this technical study might reflect current practice in the design of household survey programmes in developing countries. Consideration of current practices helps us to keep in mind the constraints imposed by developing country requirements and resources. However, the inclusion of a particular case-study does not necessarily mean that it represents the best possible design or procedure under the circumstances. Review of the case-studies as a group has suggested a number of design features where improvements are possible: these are discussed in the text.

The case-studies appear in alphabetical order, by country, and are presented in a standard format to facilitate comparisons between countries. Of the 11 case-studies, 9 are for integrated household survey programmes in developing countries. A tenth case-study, for Australia, is included to illustrate a somewhat more complex design that is being used by a country which is more developed but still has to deal with many of the environmental constraints found in developing The eleventh and final case-study describes the United countries. States master sample of agriculture, which was the first large-scale application of the master sample concept. While designed primarily to produce samples for farm surveys, the master sample of agriculture was frequently used to provide samples for surveys of households in the In addition to being of historical interest, it provides rural sector. a good illustration of the efficiency of a multistage area sample design in a situation where accurate detailed maps were readily available.

The case-studies are based mostly on the documentation that was available and is identified at the end of the description for each country. For some countries, the documentation covered designs and procedures already in use; for other countries it consisted primarily of proposals submitted by technical advisers. Therefore, it should not be assumed that the designs described in the case-studies were fully implemented in all of the countries. Drafts of the case-studies were sent for review to the organizations responsible for the survey programmes in 10 of the 11 countries (the United States case-study was excluded from this kind of review since it covered a system that was no longer active and was adequately described in published documents). The Statistical Office wishes to thank the many statistical offices that did respond and has done its best to incorporate their comments in the case-studies which follow.

CASE-STUDY: Australia

<u>Type of case-study</u>: This case-study describes the design of a master sample for use in multiple rounds of a monthly household survey and in special supplementary surveys.

Basis for case-study: This summary describes the master sampling frames (known as the Population Survey Framework) and the master samples used by the Australian Bureau of Statistics (ABS) for all household surveys conducted during intercensal periods. Australia conducts censuses of population and housing every five years. Following each census, a new master sampling frame is developed and a new master sample selected for use during the subsequent five-year period. References 1 and 2 are the ABS documents which provided the information for this case-study.

<u>Substantive requirements</u>: The ABS household survey programme consists of a Monthly Population Survey (MPS) and a series of special supplementary surveys with data collection periods ranging from 2 to 12 months. The MPS consists of a labour force survey augmented, in most months, by supplemental modules, covering a variety of topics, such as labour force activity, income, migration and education. Special supplementary surveys during the period 1979 to 1984 are listed in Exhibit 1.

The survey population for the MPS is all persons 15 and over, excluding members of the permanent defence forces, selected official personnel of overseas governments, and other overseas residents in Australia. Residents of special dwelling places, such as hotels, hospitals, prisons and construction camps, are included in the target population.

The surveys are designed to produce separate estimates by state and territory. Overall sampling fractions vary by state: they represent a compromise between proportional sampling and constant sample sizes. The stratification of the units in the master sampling frame permits the preparation of estimates for smaller political subdivisions, if desired.

<u>Operating environment and constraints</u>: The population of the Commonwealth of Australia was estimated to be 15,462,000 in July 1984. The population density is very low — 2.0 persons per square kilometer -- and much of the population is concentrated in coastal areas.

The country is divided into 6 states and 2 territories: the Northern Territory and the Australian Capital Territory. Each state is divided into metropolitan (capital city) and extra-metropolitan (rest of state) areas. Censuses of population and housing are conducted at five-year intervals.

	Dates	Number of households*	Topics
1.	FebMay 1979	12,200	Employment benefits, Working conditions, Sight, hearing and dental health
2.	SeptDec. 1979	15,500	Income Education
3.	FebMar. 1981	33,000	Disability Working hours
4.	MarApr. 1982	17,000	Families Working arrangements
5.	SeptDec. 1982	16,000	Income and housing Life insurance and pensions Trade qualifications School attendance
6.	Feb.1983-Jan.1984	17,200	Travel Health insurance Crime victimization
7.	JanDec. 1984	7,500	Household expenditure

Exhibit 1 Supplementary surveys conducted during the period 1979 to 1984

*The sample for the MPS is about 33,000 households

Household surveys are conducted by personal (face-to-face) interview. The same interviewers do the interviewing for the MPS and the supplementary surveys. Editing, coding and data entry for completed questionnaires are performed in ABS state offices.

<u>Master sampling frame</u>: Following each quinquennial census of population and housing, a master sampling frame is constructed from the census materials. The basic frame units are called census collector's districts (CDs). Maps and census dwelling unit counts are available for all CDs.

The CDs are used as PSUs in all metropolitan areas and in the more densely populated extra-metropolitan areas. In the less densely populated extra-metropolitan areas, the PSU is a group of CDs, usually comprising one or more local government areas (LGAs).

The ABS collects data on all building approvals (permits) granted in Australia, for use in its construction statistics programme. These data are also used to update the master sampling frame and to revise the master sample biannually in areas where the growth has been concentrated in certain CDs.

A separate frame is maintained for special dwellings (SDs). Frame information for each SD includes the geographic location, type of SD (hotel, hospital, prison, etc.) and number of occupants. Lists are compiled by ABS state offices from the latest census and other sources and forwarded to the central office for use in sample selection. The states update their SD lists biannually to show additions, deletions and changes (e.g., significant changes in numbers of occupants).

<u>Master sample</u>: The master sample for the ABS household surveys is a stratified sample of CDs. The areas covered by the more densely populated strata are called self-representing areas (SRAs). In the SRA strata, a single stage of sampling is used to select the master sample CDs; selection is systematic, with probability proportionate to size. The remaining, less densely populated strata are known as the "sampled strata." In these strata, selection proceeds in two stages. At the first stage, a sample of either one or two PSUs, each consisting of a group of CDs, is selected from each stratum. At the second stage, sample CDs are selected from each sample PSU. Selection with probability proportionate to size is used at both stages.

The stratification used in the design of the master sample is relatively complex. A full account is given in references 1 and 2. Exhibit 2 (adapted from reference 2) shows the basis for stratification in each of the 6 states and 2 territories. Population density is an important factor in stratification of the extra metropolitan regions and, as already mentioned, additional stages of sampling are used in the less densely populated areas.



Exhibit 2 Stratification for Selection of Master Sample PSUs

Notes

- 1. Statistical districts (group 4) are extra-metropolitan areas containing the larger cities or towns and surrounding rural areas.
- 2. Groups 4 to 7 are further subdivided into urban and rural strata.

The measures of size assigned to CDs or CD groups in each stratum are integers determined by dividing dwelling unit counts by the desired cluster size for the stratum and rounding the quotient. Desired cluster sizes are those considered to be optimum based on an analysis of variance and cost data; they vary from 4 to 10 dwelling units depending on the type of stratum. The use of these "cluster measures of size," in combination with appropriate procedures for sampling within the master sample CDs, makes possible the selection of a self-weighting sample for each state and territory.

With a few exceptions, all samples of private dwelling units for the MPS and supplementary surveys during a five-year period are selected from the initial set of master sample CDs. Exceptions occur primarily when (1) all of the blocks and clusters of dwelling units within blocks in a CD have already been used in a survey or (2) some of the CDs in a stratum have grown by unusually large amounts. In such cases, new CDs are added to the master sample.

Strictly speaking, there is no master sample of special dwelling places except in the so-called sampled strata. Complete lists of SDs are maintained for the self-representing areas. Initial samples for the MPS and supplementary surveys and replacement samples for the MPS are selected from these universe lists, which are periodically updated. In self-representing areas, the selection is entirely independent of the sampling of private dwellings, so that not all sample SDs are located in the CDs selected for the private dwelling sample. For the sampled strata, however, all initial and replacement samples of SDs are selected from lists for the master sample PSUs. Hence, the lists of SDs for these PSUs are, in effect, a master sample of SDs.

<u>Subsampling from the master sample</u>: The selection of the initial sample of private dwellings (PDs) for the MPS will be described first. This will be followed by a brief discussion of the selection of parallel samples for supplementary surveys and of replacement samples used in rotation of the MPS sample. Finally, the selection of special dwelling samples will be described.

- a. <u>Initial sample of PDs for the MPS</u>. There are five steps in the selection:
 - (1) Counting the sample CDs. Field staff locate and count all private dwellings in each sample CD. Most rural and some urban CDs are counted from the air.
 - (2) Formation of blocks in sample CDs. Dwellings in sample CDs are grouped into bounded areas, called blocks, of appropriate size. Blocks are generally expected to contain the equivalent of 4 to 8 "clusters" (see previous section on "Master sample") in urban areas and one cluster in rural areas.

- (3) Selection of blocks. Usually only one block is selected from each master sample CD. Selection is with probability proportionate to the cluster measure of size.
- (4) Listing of dwellings in sample blocks.
- (5) Selection of sample clusters. The dwellings in each sample block are listed in a prescribed order and numbered consecutively. For blocks with more than one cluster, each cluster is a systematic sample of listed dwelling units. For example, in a block with cluster measure of size equal to four, the clusters would be:

Cluster no.	Dwellings	
1	1, 5, 9, etc.	
2	2, 6, 10, etc.	
3	3, 7, 11, etc.	
4	4, 8, 12, etc.	

One of these clusters is selected with equal probability. Taking into account the selection probabilities associated with the sample CDs, the end result of these five steps is a self-weighting sample of PDs in each state and territory.

- b. <u>Parallel samples</u>. Parallel samples for supplementary surveys are also selected from the master sample of CDs. For some of these surveys all of the master sample CDs are used; for others only a subset is required. Within the master sample CDs, selection of blocks is generally limited to blocks not used for the MPS.
- c. <u>Replacement samples for the MPS</u>. The block clusters in the initial MPS sample are systematically subdivided into eight "rotation groups." After each month's survey one of these rotation groups is dropped from the sample and replaced by a new sample of block clusters of the same size. Depending on what is available (i.e., has not already been used in the MPS or a supplementary survey), the new block clusters are taken, in order of preference, from the same sample block, from a different block in the same CD, or (rarely) from a new sample CD. This order of preference minimizes the additional work needed to count dwellings and form blocks in new sample CDs and to list dwellings in new sample blocks.
- d. <u>Samples of special dwellings</u>. Each SD on the list is assigned a measure of size based on expected occupancy. SDs whose measures of size exceed four times the state sampling interval are selected with certainty; others are selected with probability proportionate to size. SDs in the sampled strata have their measures of size multiplied by appropriate factors to compensate for the sampling of PSUs. Each sample SD for the MPS is assigned to one of the eight

rotation groups. One of these groups is dropped after each month's survey and replaced by a new sample of SDs selected in the same manner.

<u>Remarks</u>: The Australian Population Survey Framework (master sampling frame) and master sample comprise an efficient, flexible facility for an IHSP. The preceding account touches only the highlights; much more can be learned by studying the detailed descriptions in the two references listed below. Some of the key features of the system are:

- A design in which the costs of counting dwellings and forming blocks in sample CDs and listing dwellings in sample blocks are spread over several surveys and survey rounds.
- o Minimization of respondent burden. At most, a dwelling can be included in the MPS in eight consecutive months during a five-year period.
- o A sample designed to produce sufficiently reliable subnational estimates when needed. The design reflects a compromise between equal reliability of estimates for states and territories vs. maximum reliability for national estimates.
- o Samples that are self-weighting at the state and territorial level.
- A conscious attempt to optimize the sample design in each type of area by varying cluster sizes and the number of stages of sampling.
- A considerable degree of decentralization in the operations required to select sample blocks and private dwellings within the master sample CDs.
- o Systematic procedures for updating the master sampling frame and the master sample during intercensal periods.

References:

Australian Bureau of Statistics.

- 1. 1983 The Population Survey Framework at the Australian Bureau of Statistics, by D.I. Cocking, Sampling Section.
- 2. 1984 Sampling Concepts and Procedures Manual.

CASE-STUDY: Botswana

Type of case-study: This case-study describes the design of a master sampling frame and a master sample of PSUs that was selected from it for use in Botswana's Continuous Household Integrated Programme of Surveys (CHIPS). Half of the master sample PSUs were used in the Primary Health Care (PHC) Evaluation Survey, which was conducted from August 1983 to May 1984. It had been planned to use the remaining PSUs later in other surveys, but this plan was dropped: samples for subsequent labour force and income and expenditure surveys were selected independently from the master sampling frame.

Basis for case-study: This summary is based primarily on reports prepared by two different technical advisers. The first adviser assisted with the design of the master sample during a mission to Botswana in April 1983 and the second adviser evaluated its selection and use during a mission in November 1984.

<u>Substantive requirements</u>: The immediate need, at the time the master sample was designed, was to design and select a sample for a survey to evaluate the need for, availability and use of primary health care facilities. The plan was to conduct the PHC Evaluation Survey in three consecutive rounds, each lasting three or four months, so that data from each round could be released separately and data from successive rounds could be cumulated.

It was expected that the PHC Evaluation Survey would be followed by either a labour force or an income and expenditure survey, and that the master sample could provide samples for those surveys.

The available documentation did not indicate whether any subnational data were wanted from the PHC Evaluation Survey or other surveys. The design that was developed appears to be aimed at minimizing sampling error for national estimates.

<u>Operating environment and constraints</u>: The population census of Botswana was conducted in 1981. Approximately 170,000 households were counted in the census. The country is divided into 12 administrative districts and has six urban centres referred to as towns.

For the PHC Evaluation survey there were seven interviewer teams, each consisting of a supervisor and three interviewers. Interviewing in each PSU was usually conducted by a single team, and a team was responsible, on the average, for covering about 30 PSUs.

<u>Master sampling frame</u>: The primary source of the frame for the master sample was the 1981 Population Census of Botswana. For the census, the country was divided into enumeration areas (EAs). Census enumerators compiled lists of the dwelling units in their EAs and recorded the number of households in each dwelling unit. The frame was divided into five primary strata: towns, villages, lands, freehold farms and cattle posts. In the town stratum, census EAs or groups of EAs were used as PSUs. Outside of the towns, census EAs were often a mixture of villages, lands, freehold farms and cattle posts. These EAs had been subdivided, for the purpose of selecting samples for current agricultural surveys, into "blocks", each of which was confined to one of the four primary strata (excluding towns). Blocks, groups of blocks or parts of blocks were used as PSUs in these four strata.

Census counts of households were available for the PSUs (EAs) in the town stratum. For the other primary strata, counts of households by block were not available, but counts of dwellings (occupied plus vacant) were available from a pre-census mapping operation.

<u>Master sample</u>: The master sample was a sample of PSUs. A separate sample of PSUs was selected from each of the five primary strata: towns, villages, lands, freehold farms and cattle posts. The method used to select PSUs was designed to facilitate the use of the master sample to obtain a self-weighting sample of dwellings for use in specific surveys. For this purpose, each selected PSU had a designated subsampling interval: either 1 (take all), or a small integer, generally 2, 3 or 4.

The sample PSUs in each stratum were assigned systematically to one of six random subgroups. Three of these subgroups, with a total of 219 PSUs, were used for the PHC Evaluation Survey. Use of this sample of PSUs, with subsampling at the rates indicated, produced a self-weighting sample with an overall sampling fraction of 1 in 16.

Three different kinds of selection procedures were used to select PSUs in the five primary strata:

(1) Town stratum. EAs or EA groups were used as PSUs. EAs with a small number of census households were combined with adjacent EAs. Each EA or EA group was given a measure of size equal to its census household count divided by 50 and rounded to the nearest integer. The EAs in the town stratum were listed by town and by EA number within the town. The sample EAs were selected systematically, with probability proportionate to their measure of size. The designated subsampling interval for each EA selected was its measure of size. For example, an EA with a household count of 220 would receive a measure of size 4 (220 - 50 and rounded) and, if selected, would be subsampled at the rate 1 in 4. This selection method was designed to result in an interviewer workload of about 50 households in each sample PSU.

(2) <u>Village stratum</u>. Blocks (parts of census EAs) or groups of blocks were used as PSUs. Blocks with fewer than 60 dwellings were combined with adjacent blocks. Blocks and block groups were stratified by size. Within each size stratum, blocks were selected for the master sample systematically, with equal probability, as follows:

Block size (dwellings)	Sampling interval	Designated within block sampling interval
60 to 119	8	l (take all)
120 to 274	4	2
275 and over	2	4

If all sample PSUs were used in a survey, the overall sampling rate for each of the block size substrata would be 1 in 8. Since only 3 of the 6 random subgroups of PSUs were used for the PHC Evaluation Survey, the overall rate was 1 in 16.

(3) Lands, freehold farms and cattle post strata. The same general method was used in all three strata. Blocks or groups of blocks were used as PSUs. Blocks with small dwelling counts were combined with geographically adjacent blocks to form PSUs. Blocks with large dwelling counts were treated as though they contained 2, 3 or 4 equal-sized PSUs. In each primary stratum, the PSUs were stratified by size. After the PSUs were ordered geographically within each of the size substrata, the PSUs selected systematically, sample were with equal probability, at the rate of 1 in 8. When a "PSU" was selected from a large block, it was designated for subsampling, using an interval equal to the number of "PSUs" in the block.

<u>Subsampling from the master sample</u>: As stated earlier, the sample PSUs in each primary stratum were allocated systematically to six subgroups and three of these were used for the PHC Evaluation Survey.

Some of the sample PSUs required further sampling in order to provide a self-weighting sample of dwellings. For each of these PSUs, the appropriate sampling interval had been designated as part of the process of selecting the master sample of PSUs.

Two different methods were used for sampling within PSUs. In the town stratum, all sampling was accomplished by selecting dwellings systematically, with a random start, from the census listings, using the sampling interval designated for the PSU. In the other four primary strata, the same method was used in some of the PSUs requiring further sampling. In other PSUs, however, the sampling was accomplished by the field supervisor, who divided the PSU into segments, equal in number to the designated sampling interval for the PSU and roughly equal in size (number of dwellings), and then selected one of these segments at random, with equal probability.

<u>Remarks</u>: For the PHC Evaluation Survey, the sample within the sample PSUs consisted entirely (possibly with minor exceptions) of dwellings listed in the 1981 census, so households living in dwellings constructed after the census were not represented. It would have been possible, if the master sample PSUs has been used for other surveys, to use field listing procedures that would ensure representation of new dwellings.

The field work for the three subgroups of PSUs was not confined to three separate four-month periods, as originally planned, so it was not possible to use the survey results for analysis of seasonal variations in health variables.

References:

1. Kalton, G.

1984 Report of mission to Botswana: 6 November - 1 December, 1984. Statistical Office, United Nations.

2. Verma, V.

1983 Report on a mission to Botswana: April 4-22, 1983. Statistical Office, United Nations.

CASE-STUDY: Ethiopia

<u>Type of Case-Study</u>: This case-study describes the use of a master sample of PSUs for multiple household and agricultural surveys in Ethiopia during the period 1980 to 1983.

Basis for Case-Study: This summary is based on a February 1983 report, The Rural Integrated Household Survey Programme, issued by the Central Statistical Office of Ethiopia.

<u>Substantive Requirements</u>: In 1980, the Central Statistical Office initiated a National Integrated Household Survey Programme (NIHSP), with assistance from FAO/UNDP and UNICEF. During the period covered, the NIHSP surveys focused mostly on agricultural activities and the socio-economic characteristics of agricultural households. Starting in 1982, non-agricultural households in rural areas were included in some surveys. Nomadic groups and urban households were excluded from coverage.

Major topics covered in surveys include: agricultural production and crop forecasts; demographic characteristics and vital rates; household income, consumption and expenditures; labour force; health and nutrition; producer and retail prices; and community variables. For most topics national and regional data were needed.

Agricultural data were collected annually. Each of the other topics was surveyed only once or twice during the 1980-1983 period; however, for some of these topics, sample households were interviewed on two or more occasions, partly to take account of seasonal variations and partly to provide improved estimates of changes during the survey period.

Details of topics covered, reporting units, collection methods and timing for the nine types of surveys conducted during 1980 to 1983 are shown in Exhibit 1.

Operating environment and constraints: Ethiopia is a large country with a land area of 1.22 million square kilometers. No population census had been conducted through 1983, but the total population as of 1979 was estimated to be 30 million. About 12 percent of the population live in urban areas.

The country has 14 administrative regions, 102 provinces (awrajas) and 600 sub-provinces (weredas). Within the sub-provinces there are about 1,300 urban dwellers' association areas and 25,000 farmers' association areas. The average size of a farmers' association area is about 250 households.

The field staff established for the NIHSP consisted of 500 interviewers (one per PSU), plus 103 field supervisors and higher level

officials. Interviewers, who are mostly high school graduates, are recruited at the regional level and are stationed in the PSUs to which they are assigned.

<u>Master sampling frame</u>: The master sampling frame for the NIHSP was a list of farmers' associations (FAs) compiled during June-July 1980. The list covered 12 of the 14 regions and within these regions it covered 419 sub-provinces in 77 of the 85 provinces in the 12 regions. No estimate was given of the proportion of the total or rural population covered by the frame.

For each of the 18,989 frame units (FAs), the frame information consisted of the name of the FA, the name of the sub-province in which it was located, and an estimate of the number of members.

Master sample: The master sample consisted of 500 FAs, which served as PSUs for subsampling to meet the requirements of each survey.

To select the master sample, the frame units were stratified by province within region. The 500 sample PSUs were allocated to the 12 regions by size (number of FAs), subject to restriction of a minimum sample of 34 PSUs from each region. Then, for each region, the allotted sample of PSUs was allocated to provinces in proportion to the number of FAs in each province. Within each province, the designated number of FAs was selected by sampling without replacement, with probability proportionate to size (number of FA members).

For the purpose of sampling within the PSUs, two field listing operations were conducted in the master sample PSUs, the first in 1980 and the second in 1982. In the 1980 listing operations, all households in the sample PSUs were listed and identified as agricultural or non-agricultural. In addition, separate listings of holders (persons in charge of agricultural holdings) living in agricultural households were prepared. In the 1982 listing operation, a fresh household listing was prepared, but there was no separate listing of holders.

<u>Subsampling from the master sample</u>: With one exception, all collection of information in the NIHSP surveys was limited to units located in the 500 master sample PSUs. The exception occurred in the Current Agricultural Survey, in which data for state farms were obtained from administrative reports. These data were then combined with sample estimates for private peasant holdings and cooperatives based on data from the master sample PSUs.

There was considerable overlap among the samples for the different surveys, as will be seen from the following descriptions of the method used in each survey to sample from the listings for the master sample PSUs. The surveys are listed in the same order as in Exhibit 1:

(1) Current Agricultural Surveys. For the first survey (Nov. 1980

 Jan. 1981), 25 holders were selected in each sample PSU by

simple random sampling without replacement from the 1980 listing of holders. The same sample holders were included in the second survey (Nov. 1981 - Jan. 1982). For the third survey (Nov. 1982 - Jan. 1983) 25 agricultural <u>households</u> were selected from the 1982 listings and all holders living in these households were included in the sample. Cooperative farms were not subsampled: in each survey all cooperative farms in the master sample PSUs were included. In each of the three surveys, a subsample of fields belonging to the sample holders was selected for objective measurement of crop areas and yields.

- (2) Crop Production Forecasting Surveys. The first 10 of the 25 holders selected for the first current agricultural survey were included in the sample used for the 1981 and 1982 crop production forecasting surveys.
- (3) Demographic Survey. A sample of 100 agricultural households was selected in each PSU, with probability proportionate to the number of holders in the household. The same sample was used for both rounds of the survey.
- (4) Income, Expenditure and Consumption Survey. The sample for each PSU consisted of 24 households associated with sample holders selected for the first current agricultural survey. The 24 households were randomly subdivided into three groups of 8 households, to be interviewed in the first, second and third months of each quarter respectively.
- (5) Labour Force Survey. The sample of households selected for the Income, Expenditure and Consumption survey was also used in this survey.
- (6) Price Data Collection. The method used to select retail outlets and producers are not specified, except to state that if the PSU had no retail outlets, price data were collected from outlets in an adjacent PSU.
- (7) Nutrition Survey. The 25 agricultural households in each PSU selected for the third current agricultural survey (using the 1982 listings) were used for the nutrition survey. In addition, up to five non-agricultural households were selected from the listing for each PSU.
- (8) Health Survey. The sample households for the nutrition survey were also used for the health survey.
- (9) Survey of Community-Level Variables. No subsampling was involved. In each of the 500 master sample PSUs, the required information was obtained by interviews with government or farm association officials.

<u>Remarks</u>: Some aspects of this design for an integrated household survey programme merit special attention:

- (1) Certain of the agricultural households in the master sample PSUs were subjected to an unusually heavy response burden during the period November 1980 through May 1982. They were interviewed twice in current agricultural surveys, four times in the Labour Force Survey, and about 35 times in the Income, Expenditure and Consumption Survey. One wonders what effect this may have had on cooperation rates and on the accuracy of responses as the interviewing progressed.
- (2) No information was provided on the treatment of sample households whose composition changed from one survey to the next, or from visit to visit within a survey. The fact that new household listings were considered necessary in 1982 suggests that changes in households were frequent enough to cause some problems.
- (3) In the third current agricultural survey, the second-stage sampling units were changed from holders to agricultural households. This was evidently done for two reasons: to simplify the field listing operation and to make it easier to integrate data from the household surveys and the surveys of holdings.

Reference:

С

1. Central Statistical Office (Ethiopia) 1983 The Rural Integrated Household Survey Programme: Methodological Report for the Period of 1980-83, Statistical Bulletin 34. Addis Ababa.

		Unite	Collection	Number of interviews/	
Survey	DUFBUID	covered	metnod	ODSELVATIONS	кепаткв
<pre>1. Current Agricultural Surveys</pre>	Nov.'80 - Jan.'81	Holdinge Fields	Interview, area measurement and crop-cutting		Repeated in '81- '82 and '82-'83
2. Crop Production Forecasting Surveys	18.'Inf	Holdings	Area measurement	I	Repeated in Jun Aug. '82
3. Demographic Survey Round 1 Round 2	Jan. '81 Jan. '82	Agricultural households	Interview		
4. Income, Consumption & Expenditure Survey	May '81 - May '82	Agricultural households	Interview	35	2 interviews per week in 4 of 12 months
5. Labour Force Survey	May '81 - May '82	Agricultural households	Interview	4	Interviews first week of each quarter
6.Price Data Collection	May '81 - May '82	Retail outlets Agricultural households	Interview Interview	12 12	Monthly interviews
7. Nutrition Survey Round 1 Round 2	Aug. '82 Jan. '83	Rural households	Interview	г	Interview included anthropometric meas- urement of children
8. Health Survey Round 1 Round 2	Аи g. '82 Јап. '83	Rural households	Interview		
9. Survey of Community- Level Variables	Jun Aug.'81	Local officials	Interview	1	

Exhibit 1 NIHSP Surveys, 1980-83

CASE-STUDY: India

Type of case-study: This case-study describes the design of a master sample for use in conducting multiple surveys during a single year.

<u>Basis for case-study</u>: This summary describes the relevant aspects of the design of the Indian National Sample Survey (NSS) in the mid-1970s, as described in a 1975 paper by Rao and Sastry (Reference 2). Some details on the use of materials and results from decennial censuses to develop master sampling frames for the NSS come from a paper by Murthy (Reference 1).

<u>Substantive requirements</u>: The multiyear plan for the NSS described in Reference 2 called for collection of data on many topics. These topics were combined into five major groups:

- (1) Employment, unemployment, rural labour and consumer expenditure.
- (2) Self-employment in the non-agricultural sector.
- (3) Population, births, deaths, disability, morbidity, fertility, maternity, child care and family planning.
- (4) Debt, investment and capital formation.
- (5) Landholding and livestock enterprises.

Over a ten-year period, the first two groups were to be covered twice, at five-year intervals, and each of the last three was to be covered only in one year. This accounted for seven of the ten years, leaving three years open for other topics of current interest.

Earlier, the NSS had included a series of "land utilization surveys" and "crop cutting experiments" to provide estimates of crop areas and yields that were independent of the "official" administrative estimates. Starting in 1973-1974, however, NSS activity in this area was limited to an independent sample check on the field work of the administrative agency.

For each year, or round, of the NSS, the period of data collection is divided into equal segments of 2 to 3 months, called subrounds. The samples used are large enough to provide separate estimates for each subround.

The primary objective of the NSS is to provide reliable national and state estimates (India has 22 states and 9 union territories). However, an increasing demand developed for substate estimates, and the NSS has met this demand by establishing parallel central and state samples in each state, with the states using their own resources to conduct the field work for the state samples. The increase in sample size is sufficient to provide separate estimates by region (groups of administrative districts) within each state.

Operating environment and constraints: The population of the Republic of India was estimated to be 746,388,000 in July 1984, with a density of 227 persons per square kilometre. In the mid-1970's, about 80 percent of the population lived in rural areas and 70 percent were illiterate. Many different languages are spoken.

The country is divided into 22 states and 9 union territories. States are subdivided into administrative districts and sub-districts. A decennial population census is conducted regularly in years ending in 1, e.g., 1971.

A permanent field force was established for the NSS. Those in charge of the survey feel strongly that only a permanent staff of well-trained investigators can produce reliable results, given the difficult conditions under which the surveys are conducted and the variety of topics covered. As of 1975, the NSS field staff consisted of 1,307 assistant superintendents and 253 1,254 investigators, superintendents. One full-time supervisor guides the work of a team of four investigators. The work was organized under 43 regional and 123 local offices.

<u>Master sampling frame</u>: The frame was constructed from the 1971 decennial population census. Small area units with defined boundaries that had been established as enumeration blocks for the census were used as frame units and as PSUs for the NSS. These units were called census villages in rural areas and census blocks in urban areas. The census population count was available for each frame unit, as well as identification of the political subdivisions in which it was located. No definitive information is given on the size distribution of frame units; however, the author of reference 1 recommended that enumeration blocks formed for the census in both rural and urban areas have a population of 600 to 800 persons.

Reference 2 indicates that there are frequent changes in the boundaries of census blocks; therefore the urban block definitions are updated periodically during the inter-censal period.

<u>Master sample</u>: A new sample of PSUs (census blocks and villages) is selected for each annual round of the NSS. The sample of PSUs for each round is a master sample in the sense that the household listings prepared for the sample PSUs are used as a secondary sampling frame for separate samples selected for surveys on different topics and for subrounds of specific surveys.

Primary stratification of PSUs is by type of area (rural or urban) and, for rural strata, by regions within state. Secondary strata in rural areas are formed to be about equal in size. They are compact geographical areas which, as far as possible, are homogeneous in terms of population density and cropping patterns. Secondary stratification in urban areas is based primarily on population size of towns and villages.

Allocation of sample PSUs to states is made separately for rural and urban areas. For rural areas, the allocation is based on two variables, rural population and area under food crops, subject to a minimum number of rural PSUs in the sample for each state. For urban areas, the allocation is based solely on urban population, also subject to a minimum number of urban sample PSUs in each state. Within states, rural sample PSUs are distributed equally among the secondary strata and urban sample PSUs are allocated to secondary strata in proportion to their population. For the 1974-75 round of NSS, the central sample of PSUs consisted of 8,512 villages and 4,872 urban blocks. A parallel sample of the same size was selected for use by the states.

Starting with the 28th round of the NSS (1973-74), PSUs within secondary strata were selected with probability proportionate to population, with replacement (to enable easy computation of sampling errors).

The second-stage sampling units are households for socio-economic surveys and clusters of fields for crop surveys. Field investigators prepare listings of these units in the sample PSUs, including any auxiliary information that may be needed to improve the efficiency of the second-stage sample selection.

<u>Subsampling from the master sample</u>: In earlier periods, the NSS was completely integrated in the sense of collecting data from the same sample households on more than one broad topic. However, this approach was abandoned, mostly because of respondent fatigue. As of the time covered by this report, separate samples of households (or clusters of fields for crop surveys) were selected from the master sample PSUs for each survey.

The usual practice for selecting second-stage units in a sample village or urban block requires that the listed units first be arranged in some order determined on the basis of relevant auxiliary information obtained during the listing. Then sample units are chosen systematically with a random start and a sampling interval calculated so as to make the design self-weighting at the state level.

When a topic is to be covered in more than one sub-round, as is usually the case, the sample units are divided equally among the sub-rounds at the PSU level. The NSS design also uses inter-penetrating subsamples for both the central and state samples. Each subsample is selected according to the scheme described above and each is capable of providirg valid sample estimates. Separate estimates are made for each subsample. This procedure serves as a broad check on the survey results and provides an indication of the overall variability of sample estimates, including the effects of variable non-sampling errors.

<u>Remarks</u>: Reference 2 discusses the gains achieved by conducting more than one survey in the same set of sample PSUs. Travel, listing and other overhead costs associated with each PSU are high, and this design permits these costs to be shared by more than one survey. The number of sample PSUs for each survey is larger than it could be if an independent sample of PSUs were selected for each survey, thus reducing the between PSU component of the sampling variance. The design permits investigators to remain in each sample PSU long enough to establish rapport with the respondents.

References:

208-220.

- Murthy, M.N. 1969 Population census as the source of sampling frame in India. Sankhya, 31 (8): 1-12.
- 2. Rao, V.R. and Sastry, N.S. 1975 Evolution of total survey design: the Indian Experience. <u>Bulletin of the International Statistical Institute</u>, 46(1):

CASE-STUDY: Jordan

<u>Type of case-study</u>: This case-study describes the use of a master sample for multiple rounds of a multi-purpose household survey.

<u>Basis for case-study</u>: The case-study describes the proposed design of a master sample for use by the Department of Statistics of the Hashemite Kingdom of Jordan in a multi-purpose household survey (MPHHS). The information presented in this case-study is based on the report of a technical adviser who worked with the Department of Statistics in June-July 1980 (Reference 1). For several of the design features, the adviser's report provided two options, giving the pros and cons of each and identifying a preferred option. This case-study describes the design as though the preferred options had been adopted in each case.

Substantive requirements: The survey population is the noninstitutional population of the country, excluding the occupied West Bank, nomads living in remote areas and persons living in hotels. The proposal does not mention any requirements for subnational estimates.

The plan called for the master sample to be used for a period of five years, from 1981 to 1985. Two survey rounds, one in April and one in October, were scheduled for 1981 and 1982. For the three remaining years, four rounds per year, in January, April, July and October, were planned.

The initial schedule for coverage of topics was as follows:

Topic

Coverage

All rounds
All rounds
April round, annually
April round, annually ¹
July round, annually ²
October round, annually
October round, annually
January and July rounds, 1983
July round, 1984
All rounds, 1985

¹ Starting in 1982 ² Starting in 1983

A later communication indicated that not all of these topics were covered as planned. Core items were covered in all rounds, manpower in the 1981 and 1982 rounds and vital statistics in the October 1981 round. Internal migration was scheduled to be covered in the May 1986 round.
Operating environment and constraints: The count of households in the 1979 population census of Jordan was 320,248. The estimated population as of July 1984 was 2,545,200, with a density of 28.5 persons per square kilometre. The majority of the population is concentrated in Amman, the capital city. The country is divided into five administrative divisions called governorates.

<u>Master sampling frame</u>: The 1979 census of population was the main source for the master sampling frame. Census data were collected for 41 urban and 966 rural localities. Urban localities were those with a population of 5,000 or more.

For the census, some urban localities were subdivided into sectors, units and blocks. Other units and all rural localities were considered to be a single sector and unit and were subdivided into blocks. Census counts and listings of households and structures were available by block. Census structure numbers were stencilled on to the existing structures during the census.

The census materials were organized to provide a frame from which the master sample could be selected. For the urban frame, census blocks or groups of blocks were used as PSUs. Blocks having census household counts of about 10 were combined with nearby blocks in the same unit to form PSUs of about 20 households. The urban PSUs (blocks and block groups) were listed by locality, starting with the largest locality and proceeding in decreasing order of population. Within each locality, the units were listed in an order based on a rough measure of socio-economic status (SES). The SES ordering patterns for units were from low to high in odd-numbered localities and from high to low in even-numbered localities.

For the rural frame, localities or groups of localities were used as PSUs. Localities with census household counts of about 10 were combined with nearby localities. The rural PSUs were listed in order by size stratum and by governorate within size stratum. Four size strata, based on census population counts, were used: under 500; 500 to 999; 1,000 to 2,999; and 3,000 to 4,999.

The distinctive features of the urban and rural frames are shown below:

Feature	Primary stratum Urban	<u>Rural</u>
PS Us	Blocks and block groups	Localities and groups of localities
Ordering criteria Primary	Localities in decreasing order of size	Population size class
Secondary	Units by socio- economic status	Governorate

<u>Master sample</u>: The proposal called for selection of 21 independent replicates, each containing 50 PSUs, with an average "take" of 20 sample households per PSU, for a total of about 1,000 households per replicate. The sample was to be selected in two stages in urban areas and three stages in rural areas and was designed to be self-weighting, with the same overall sampling fraction for urban and rural areas. It was determined that using an overall sampling fraction of 1 in 312 for a single replicate would result in a sample of the desired size. Based on census population counts, it was decided that 35 urban and 15 rural PSUs should be selected for each replicate.

In urban areas, a preliminary measure of size (M_1) was to be assigned to each PSU by dividing its census household count by 20 (the desired sample take per PSU) and rounding to the nearest integer. These preliminary measures were then to be adjusted to make the total for all urban PSUs exactly equal to 10,920 (the product of 35 times the sampling interval, 312). The sample of 35 PSUs for each replicate was to be selected systematically, with probability proportionate to the adjusted measures of size, using a sampling interval of 312 and an independently selected random start for each of the 21 replicates.

Sampling within sample PSUs in urban areas was to be at the rate 1 in M_1 . Thus, for sample PSUs with M_1 equal to one, no further sampling was required. For PSUs with M_1 greater than one, the basic procedure was to sample housing units from updated census listings systematically with equal probability, using M_1 as the sampling interval, with a random start. All households associated with the selected housing units were to be included in the sample.

The procedures for selecting and sampling within rural PSUs were similar to those used for urban PSUs, except that one additional stage of sampling was called for. A systematic sample of rural PSUs was to be selected for each replicate, with probability proportionate to measures of size M_1 . Then, for each selected PSU, a listing of census blocks was to be prepared, forming block groups where necessary to avoid the selection of blocks with about 10 households. The block groups were assigned measures of size M_{1j} , which were adjusted to add to M_1 (the measure of the selected PSU). Then one block or block group was to be selected with probability proportionate to the measure of size M_{1j} . For selected blocks and block groups, the subsequent sampling procedures were to be the same as for urban areas, except that the indicated sampling rate would be 1 in M_{1j} , rather than 1 in M_1 .

In summary, the master sample would consist of 21 independently selected replicates, each containing 35 urban and 15 rural blocks or block groups. Each sample block or block group would have associated subsampling instructions, indicating either that all households should be included in the sample or that a specified integer sampling interval should be used to select a sample of households. The expected number of households to be included in the sample for each replicate was 1,000. Subsampling from the master sample: The subsampling procedure would consist of selecting one or more replicates for each survey round. If necessary, the master sample could also be used for surveys other than the MPHHS.

The proposal does not include a specific plan for the selection of replicates for each round, but provides the following general criteria:

- o Avoid problems of excessive respondent burden and fatigue that would result from continued use of the same replicates in several successive rounds of the MPHHS.
- o Have some sample overlap between rounds when measurement of change is important.
- o Assign available replicates at random rather than purposively.

<u>Remarks</u>: Following are some aspects of this proposed design that are of particular interest:

- (1) The 21 replicates were to be selected independently, i.e., using the same sample selection procedures and intervals at each stage, but independently selected random starts. With this procedure it would be possible to have the same set of urban or rural PSUs in more than one replicate. If this happened, it would not necessarily mean that the samples of housing units from this set of PSUs would be identical for these replicates. The selection of housing units would be done independently for each replicate and might or might not lead to the same sample.
- (2) There was evidently no interest in being able to produce separate estimates by governorate (the first-level administrative division). No minimum sample size was required for a governorate and the ordering of PSUs in the master sampling frame was such that there was no control over the proportion of urban or rural PSUs to be selected from each governorate.
- (3) The proposed design is self-weighting and uses integer sampling intervals at each stage. The latter feature is achieved by assigning integer measures of size to each primary and secondary sampling unit, and by adjusting preliminary measures of size to make them add exactly to a predetermined total.
- (4) The proposal calls for updating of the housing unit listing for each sample block prior to each round in which the block is to be used.

(5) The proposal notes that households sometimes occupy more than one housing unit. Assuming that the housing units in a sample block have been listed and numbered consecutively, the proposal recommends that such a household be uniquely associated for sampling purposes with the lowest numbered housing unit that it occupies. For example, if a household occupied housing units 10 and 13 in a sample block, it would be interviewed only if housing unit 10 had been selected.

Reference:

1. Kalsbeek, W. 1980. 23-July 1980 Trip report, Amman, Jordan, June 7, for Population International Program of Laboratories Statistics, University of North Carolina.

CASE-STUDY: Morocco

Type of case-study: This case study describes a master sample designed for use in several different surveys.

Basis for case-study: The description is based primarily on the report of a technical adviser to the Moroccan Directorate of Statistics, covering a May 1985 mission (reference 3) and updated on the basis of a communication received from the Government (reference 1).

<u>Substantive requirements</u>: The master sample was designed to accommodate all household surveys to be conducted by the Directorate of Moroccan Statistics during the intercensal period, 1982-1992. Reference 2 describes a plan for five separate surveys, as follows:

)

Topic	Number of times to be surveyed during intercensal period
Consumption and expenditures	2
Employment	10 (continuous
Demographic characteristics	3
Household industries	1
Basic needs	1

A subsequent report indicates that a continuing quarterly employment survey began in urban areas in the second quarter of 1984 and was scheduled to begin in rural areas in the second quarter of 1986 (reference 3).

For most surveys, separate data are desired for urban and rural areas in each of seven economic regions of Morocco.

Operating environment and constraints: The Kingdom of Morocco had an estimated population of 25,565,000 in July 1984, with a population density of 62.5 persons per square kilometre. About 50 percent of Morocco's rural area, containing 10 percent of the rural population, is sparsely-populated desert area. Reference 2 indicated that a separate master sample was to be designed for that part of the country. The country is divided into 39 provinces and 2 prefectures (Rabat-Sale and Casablanca). For statistical purposes, these political divisions have been grouped to form 7 economic regions. A census of population and housing was conducted in 1982. Population counts from the census were available by census district for urban areas and by commune for rural areas.

<u>Master sampling frame</u>: Different procedures were used to establish frames for the urban and rural parts of the country. For urban areas, PSUs consisting of 3 or 4 census districts (CDs) were formed, with an average of 600 households per PSU (based on census counts). Each PSU consisted of CDs from one of eight housing types: luxurious, modern, New Medina, Old Medina, industrial area, shantytown, urban <u>duoar</u> and small urban centres. The total number of urban PSUs was 2,574.

For rural areas, the actual frame was a list of communes with a specified number of "virtual PSUs" assigned to each commune. The number of virtual PSUs for a commune was determined by dividing its census count of households by 1,000 and rounding to the nearest whole number. Conceptually, the rural frame was considered to be a listing of the virtual PSUs, of which there were 1,839.

In both the urban and rural frames the primary stratification of the PSUs was by economic region. In the urban frame, secondary stratification was by the eight housing types. In the rural frame, secondary stratification was by province.

<u>Master sample</u>: In the urban areas, a one-stage sample of 536 PSUs (CD groups) was selected. The sample of PSUs was allocated to the secondary strata in proportion to their numbers of PSUs. The indicated number of PSUs was obtained from each secondary stratum by systematic selection, with probability proportionate to census household counts.

In the rural areas, a sample of 432 of the virtual PSUs was selected in two stages. The sample of PSUs was first allocated to the 30 provinces in proportion to their numbers of PSUs. Within each province, communes were selected systematically with probability proportionate to their census household counts. Under this procedure, some of the larger communes were designated for selection of two or more PSUs in the second stage of selection.

Prior to the second stage of selection in the rural areas, a field operation was carried out to divide each of the sample communes into area segments containing, on the average, 1,000 households. Then, the indicated number of segments was selected from each sample commune by systematic selection, with probability proportionate to census household counts.

Although the rural selection of PSUs actually proceeded in two stages, it was considered to be conceptually equivalent to a one-stage selection of area segments. <u>Subsampling from the master sample</u>: Subsampling procedures may be expected to vary from survey to survey. For the quarterly employment survey, the master sample of PSUs has been divided into 4 random subgroups: one of these is used each quarter. Urban sample PSUs have been divided into area segments with an average size of 50 households: these segments are used as secondary sampling units. Rural sample PSUs have also been segmented, but the average size of rural segments is twice as large -- 100 households.

These secondary sample units (SSUs) may be used in various ways, depending on the requirements of each survey. One or more SSUs could be selected from each master sample PSU included in a survey or survey round. Selected SSUs could be sampled further either by division into smaller segments or by listing and sampling of households or housing units.

For the continuing employment survey, partial rotation of sample SSUs within the sample PSUs at the rate of one-third sample replacement per year has been proposed (reference 3).

<u>Remarks</u>: Reduction in the cost of field-work for the segmentation of urban PSUs and rural communes was one of the major benefits of this master sample design. In the urban areas, segmentation was required only for the sample PSUs, which were about 25 percent of the total. The total number of rural communes is not given in reference 3, but from other sources it would appear that segmentation was required for roughly 60 percent of the communes.

References:

- Direction de la Statistique 1986 Percu sur le plan de sondage de l'enchantillon-maitre Marocain. Ministry of Planning, Kingdom of Morocco.
- 2. Garrett, J. and Stanecki, K. 1983 A master sample approach to the Moroccan intercensal survey program. <u>American Statistical Association Proceedings, Social</u> <u>Statistics Section</u>: 243-248.
- 3. Roy, G.
 - 1985 Rapport d'une mission effectuee a la Direction de la Statistique du Maroc (Rabat) du 15 au 24 Mai 1985. Institut National de la Statistique et des Etudes Economiques, Paris.

CASE-STUDY: Nigeria

Type of case-study: This case-study describes the design of a master sample for use in multiple surveys.

Basis for case-study: The master sample described in this case-study was designed for use in Nigeria's National Integrated Sample of Households (NISH) during the period April 1981 to March 1986. The description is taken from the report of a U.S. Census Bureau adviser to Nigeria's Federal Office of Statistics (Reference 1). The adviser's report described the current NISH design and included proposals for the five-year period starting in April 1986.

<u>Substantive requirements</u>: The chart below shows the surveys conducted during the five-year period from April 1981 to March 1986 (survey years run from April of one year to March of the following year):

Name of survey	Topics covered	Conducted in years:
General Household Survey (GHS)	Demographic and socio-economic characteristics	1 to 5
National Consumer Survey (NCS)	Household consumption and expenditures	1 to 5
Rural Agricultural Sample Survey (RASS)	Farm characteristics, crop area and production, livestock	1 to 5
Health and Nutrition Status Survey (HANSS)	Health and nutrition status of population	3
Labour Force Survey (LFS)	Employment, unemployment and underemployment	3 to 5
Survey of Housing Status	Supply, demand and characteristics of housing	5

The design of the Survey of Housing Status was not described in the adviser's report and is therefore not covered in this case-study.

The surveys were designed to provide national and state estimates, and separate estimates for the urban and rural areas of each state. It was also considered desirable, although not feasible with the design used, to obtain separate estimates for each state capital. In the Rural Agricultural Sample Survey (RASS), data were collected from a fixed sample of farm households throughout the survey year. In the other two continuing surveys, data were collected monthly throughout the year, but using a different sample of households each month. A similar data collection pattern was used for the Health and Nutrition Status Survey (HANSS). The interviews for the first Labour Force Survey (LFS) were conducted in December 1983. Additional surveys were carried out in December 1984 and in June and December 1985.

<u>Operating environment and constraints</u>: The Federal Republic of Nigeria had an estimated population of 98,148,000 as of July 1984 and a population density of 95 persons per square kilometre. The country is divided into 19 states. Literacy is estimated at 25 to 30 percent.

A census was taken in 1973; however, the results were discarded and population counts by enumeration area (EA) are not available.

There was a permanent field staff in each state, with up to 60 rural enumerators and 6 to 8 urban enumerators assigned to listing and interviewing for NISH. Enumerators generally worked in pairs (teams), with a field supervisor overseeing, on the average, three teams. Rural enumerators were stationed for the full survey year in a single EA; urban enumerators worked in different EAs each month.

Master sampling frame: The units of the master sampling frame were the EAs created for the 1973 census of population. The EAs were defined on sketch maps, except for an unspecified proportion of EAs for which no sketch maps were available when the frame was established. EAs were classified as urban or rural, urban EAs being those in "localities" with a population of 20,000 or more.

Since the 1973 census results had been discarded, no population counts or other data were available by EA. Based on listings of a sample of census EAs for NISH, it was found that the average population was about 600 persons for urban EAs and about 450 persons for rural EAs.

<u>Master sample</u>: Since a sample of about the same size and design was selected for each of the 19 states, it will be convenient here to describe the design of the master sample of EAs and the subsequent subsampling procedures for a single state.

Census EAs were grouped into four strata: state capital, large towns, other towns and rural, the first three all being classified as urban. Within each stratum the EAs were ordered geographically.

A double sampling procedure was used to select the master sample of EAs. In phase I, a sample of 200 census EAs was to be selected, allocated to the four strata as follows:

State ca	pital	45
Large an	d other towns	69
Rural		86
	Total	200

For the state capital and rural strata, the phase I selection was made in a single stage, i.e., the census EAs were the PSUs. In both of these strata, the designated numbers of sample EAs were selected systematically, with equal probability.

For the other two urban strata, towns were used as PSUs (except in states with fewer than four towns). A sample of three towns (either one or two from each stratum) was selected with probability proportional to the number of EAs in each town. From each of the three sample towns, 23 sample EAs were selected systematically, with equal probability.

Following the phase I sample selection, field listings were prepared for most of the 200 sample EAs. All households were listed, along with their basic attributes, including information necessary to identify farm households.

In the phase II selection, nine subsamples, each consisting of eight urban and six rural EAs, were selected from the phase I sample. In the urban strata, EAs were selected with probability proportionate to the total number of households that had been listed. In the rural stratum, EAs were selected with probability proportionate to the number of farm households listed. For any EAs not listed (generally because sketch maps were missing), the average measure of size for the stratum was used to determine the probability of selection.

Reference 1 does not give a detailed explanation of the method of selecting the nine subsamples. Since they were all selected from the same phase I sample of census EAs, they are not replicates in the strict sense and should probably be called random subgroups or pseudo-replicates. These nine random subgroups constituted the master sample for NISH during the period April 1981 to March 1986.

<u>Subsampling from the master sample</u>: The NISH sample of EAs for the first survey year (1981-1982) consisted of five of the nine random subgroups in the master sample, for a total of 40 urban EAs and 30 rural EAs in each state. In each succeeding survey year, one of the five random subgroups was replaced by a new one.

For the GHS and the NCS, the sample EAs in the urban and rural strata were allocated at random to the 12 months of the survey year, so that either 3 or 4 urban EAs and 2 or 3 rural EAs were assigned to each month.

The enumerators prepared household listings in sample EAs in the month preceding the survey month. Using sampling intervals provided to them, they selected samples for the three continuing surveys. The intervals were chosen to provide roughly the following number of sample households for the three continuing surveys:

GHS	- 20 households		
NCS	- 5 households		
RASS _	- 20 farm households		

The GHS sample of households in each EA was treated as the master sample; other surveys used either all or a subsample of these households. The selection for the RASS is carried out only in rural sample EAs.

For the HANSS, in the rural stratum, only the EAs assigned to the first quarter of the 1983-84 survey year were included, i.e., about 7 or 8 rural EAs. A sample of about 20 households was selected from each sample EA. This sample was divided randomly into four groups of five households and one of these groups was interviewed in the HANSS during each quarter of the survey year. In the urban strata, the first quarter EAs were treated the same way as in the rural stratum, but, in addition, EAs for the other quarters were covered in the months to which they had been allocated.

The sample for the first LFS included only the NISH sample EAs for December 1983, either three or four urban EAs and two or three rural EAs. In each of these EAs a sample of about 20 households was interviewed. The December 1984 and December 1985 surveys each used the October, November and December EAs for the corresponding year, and the June 1985 survey used the April, May and June 1985 EAs. Thus, the sample for each of these three surveys was about three times the size of the sample used for the first LFS in December 1983.

With the exception of the sample for the RASS, none of the samples of households was self-weighting at the state level, but all samples were designed to be self-weighting within strata.

<u>Remarks</u>: The adviser's report (Reference 1) contained several findings concerning the sample design for the 1981-1986 NISH and included some suggested new features for use in the redesign covering the 1986-1991 period. Some points of general interest are:

1. The design oversampled the urban strata. Over half of the sample EAs were urban, although only about 12 percent of the total population was urban. This was done intentionally because separate estimates were desired for state capitals, because socio-economic characteristics are believed to be more variable in urban areas and because collection costs are quite a bit lower in urban areas.

- 2. Farm households in census EAs classified as urban were excluded from the sampling frame for the RASS. The adviser recommended that the proportion of farm households thus excluded from the RASS be estimated from the listings in order to decide whether steps were needed to include them in the RASS.
- 3. Another agency of the Nigerian Government, the National Population Bureau, had a programme underway to update the frame of census EAs for use in the next population census. The adviser recommended that the two agencies collaborate in the updating operation and that the work be done in a manner that would permit use of this frame to select a new master sample for NISH as soon as possible.
- 4. The household was used as the ultimate sampling unit for NISH. The adviser noted that the reasons "household moved" or "household not found" were often given to explain why an interview was not completed for a sample household. He recommended that consideration be given to the use of either housing units or compact area clusters as the ultimate sampling units.
- 5. For the GHS and the NCS, households in each sample EA were interviewed only during the assigned month, but for the RASS, data were collected from sample farm households throughout the survey year. For this reason, and also to collect rural price data, an enumerator team was stationed in each rural sample EA for the entire year. The workload for these teams was very unevenly distributed, with the peak load occurring in the month when interviews for the GHS and the NCS, as well as the RASS, were scheduled. The report suggested some changes designed to remedy this imbalance.

Reference:

- 1. Megill, D.
 - 1985 Preliminary recommendations for redesigning the sample for the National Integrated Sample of Households (NISH). U.S. Bureau of the Census, unpublished report.

CASE-STUDY: Saudi Arabia

Type of case-study: This case-study describes the design of a sample for use in multiple rounds of a single survey.

<u>Basis for case-study</u>: This description is based on a report of the sample design for Phase IV of Saudi Arabia's Multipurpose Household Survey (MHS), covering a five-year period starting in 1981 (year 1401 in the Moslem calendar).

<u>Substantive requirements</u>: The objectives of Phase IV of the MHS were to measure changes and trends of the population over the five-year period and to collect data on the labour force, vital events, vocational training, health, nutrition, expenditures, housing and work experience. Kingdom-wide estimates were needed separately for metropolitan, urban and rural areas, as well as separate estimates for each of the country's five regions. Data were to be collected quarterly on one or more topics.

Operating environment and constraints: The estimated population of the Kingdom of Saudi Arabia was 10,800,000 as of July 1984 (Reference 3). The population density is low, about 4.6 persons per square kilometre. The country is divided into five geographic regions: Central, Northern, Western, Southern and Eastern. Each region consists of one or more administrative areas, and these are subdivided into emirates.

A census of population had been conducted in 1974, and the maps and small area population counts from the census were available for designing the Phase IV MHS.

<u>Master sampling frame</u>: The frame was developed using the 1974 census of population. The basic frame units were the emirates, which were classified as metropolitan, urban or rural, according to the following criteria:

Metropolitan: all emirates which

- (1) contained a municipality of 50,000 or more settled population, or
- (2) were economically linked to a municipality of 50,000 or more settled population, or
- (3) contained two or more municipalities of 30,000 or more settled population within 25 miles of each other.
- Urban: all emirates containing one or more municipalities of 5,000 or more settled population and not classified as metropolitan.

Rural: all other emirates.

Note: This study is based primarily on two Government documents cited as references. It is not known whether the sampling scheme envisaged in those documents has been implemented in that form.

In the metropolitan and urban areas, the emirates served as PSUs. In the rural areas, the PSUs were emirates or groups of emirates. Adjacent emirates were combined to form PSUs when necessary to ensure a minimum population of 5,000 per PSU.

<u>Master sample</u>: Within each region, the PSUs were grouped to form metropolitan, urban and rural strata. Each of the metropolitan areas, ten in all, was considered to be a separate stratum. Within each region, one urban and one rural stratum were formed, except in the Southern region, where two rural strata were formed. In all, there were 21 strata: ten metropolitan, five urban and six rural. One PSU was selected from each stratum with probability proportionate to the 1974 settled population count.

Each of the 21 sample PSUs was divided into smaller areas, called segments, with expected size in the range from 100 to 200 households. The segments were formed by the Mapping and Printing Section of the Central Department of Statistics, using maps and field visits to the PSUs. A sample of one or more segments was selected from each PSU, taking into account the PSU selection probabilities, to produce a self-weighting sample of segments of the desired overall size. The number of segments is not given in the reference, but on the basis of the estimated total sample size, there must have been roughly 300 segments selected from the 21 sample PSUs.

A "prelisting" operation was carried out in each sample segment. All structures were enumerated and all housing units and households identified. Following the prelisting, each household in a sample segment was randomly assigned to one of eight panels.

Subsampling from the master sample: For the first year of the Phase IV MHS, panels 1 to 4 were to be included in the sample. At the beginning of the second year, panel 1 would be replaced by panel 5, and so on, as shown in the following diagram (the x's indicate the years that each panel is in the sample):

Survey Year	Panel							
	1	2	3	4	5	6	7	8
1	x	х	х	x				
2		Х	Х	х	X			
3			X	X	X	Х		
4				X	Х	X	X	
5					X	X	X	Х

Thus, for each quarter, there would be a 75-percent overlap with the same quarter a year earlier. There would be 100-percent overlap between

quarters within a year, but only a 75 percent overlap between the last quarter of one year and the first quarter of the next. Households in panels 4 and 5 would be interviewed in each of 16 successive quarters, while the total number of interviews for the other panels would be smaller.

<u>Remarks</u>: The number of PSUs in the Phase IV MHS is much smaller than in most national household surveys, and the average size of the ultimate clusters (from 50 to 100 households in a segment of 100 to 200 households) is large. The reference does not explain the organization of the field staff, so a detailed analysis of the factors that led to this particular design is not possible. With this design one would expect the regional estimates, in particular, to have large sampling variances, because of the contribution of the between PSU variance component.

Other matters not covered in the reference are the treatment of nomadic population and what provisions, if any, were made for updating household listings during the five-year period for which the sample was to be used.

Reference:

- Central Department of Statistics

 n.d. The Survey Design for the Saudi Arabian Multipurpose Survey
 Phase III, 1399 A.H./1979 A.D., Ministry of Finance and
 National Economy, Kingdom of Saudi Arabia.
- n.d. The Sample Design of the Phase IV Multipurpose Household Survey of the Kingdom of Saudi Arabia. Appendix 3 of unidentified document.
- Population Reference Bureau, Inc. 1984 World Population Data Sheet (source of July 1984 population estimate).

CASE-STUDY: Sri Lanka

Type of case-study: This case-study describes the design of a master sample for use in successive annual household surveys.

Basis for case-study: A master sampling frame was developed by the Sri Lankan Department of Census and Statistics (DCS) for use in a series of household surveys conducted under an NHSCP project, starting in 1984. In addition, a master sample was designed for use in the second and third surveys in the series. The information presented in this case-study is based primarily on reports by two advisers to the survey project (see References 1 to 4).

Substantive requirements: The project plan called for three surveys:

Interviewing period	Name of survey
April 1984 - March 1985	Survey of Household Economic Activities.
April 1985 - March 1986	Labour Force and Socio- Economic Survey.
April 1986 - March 1987	Socio-Demographic Survey

The target population for the first of these three surveys consisted of households engaged in own-account economic activities. A special sample design was developed to obtain an adequate sample of this population. The 1985-86 survey was used to collect detailed data on labour force household income and expenditures. The 1986-87 survey would cover topics such as housing, health, disability, fertility and child survival. Following the first three surveys, it is expected that additional surveys will be undertaken on a more-or-less annual basis.

Each of the surveys is conducted in 12 monthly rounds, with no overlap of sample households or housing units between rounds.

The target population for the second and third surveys is the non-institutional population of the country. All three surveys were designed to produce separate estimates for each of Sri Lanka's 25 districts and for a few large cities. At the national level, separate estimates are needed for the urban, rural and estate (large holding) sectors.

<u>Operating environment and constraints</u>: The Democratic Socialist Republic of Sri Lanka had an estimated population of 15,925,000 as of July 1984, with a population density of 243 persons per square kilometre. About 21.5 percent of the population live in urban areas. The major languages spoken are Sinhala and Tamil.

The country is divided into 25 districts. The urban sector consists of municipal, urban and town council areas, which are subdivided into wards. The rural sector in each district is divided hierarchically into assistant government agent (AGA) divisions, grama sevaka (GS) divisions and villages. Areas coming under the Mahaweli development project are administered separately: the major units in these areas are designated "systems" (System A, System B, etc.).

Population censuses are normally conducted in years ending in 1, the most recent being in 1981. A mid-decade "sample census" has been proposed for 1986.

The DCS has a permanent office in each district staffed by a statistical officer and statistical investigators at the rate of about one per AGA division. The statistical investigators are responsible for field work in all DCS programmes. About 250 of them were engaged in field work for the 1985-86 Labour Force and Socio-Economic Survey.

<u>Master sampling frame</u>: The master sampling frame was developed from the 1981 Census of Population and Housing, with some updating prior to the sample selection for the 1985-86 survey. The basic frame units are the census enumeration areas, known as census blocks. Census blocks were established within wards in urban areas and within villages in rural areas. Blocks were meant to include from 70 to 100 housing units in urban areas and from 50 to 80 units in rural areas. For each block there was available a census listing form, which included a listing of the housing units at the time of the census and a sketch of the block showing the location of the housing units in relation to streets, roads and other physical features.

A listing of census blocks was available in the form of a computer printout. Sample selection and updating operations were done clerically, using copies of the census block list. The blocks were listed by district, and by sector (urban, rural, estate) within district. For the urban sector the blocks in each district were listed in order by municipal/ urban/town council and wards. For the rural sector they were listed in order by AGA division, GS division and village, and for the estate sector, they were listed by AGA division and estate. The listing included the 1981 Census counts of housing units and population for each block.

Special blocks were established in the 1981 Census for institutional and transient population. These blocks, which were included in the listing and could be identified by their block numbers, were passed over in sample selection for the household surveys. The PSUs used for the household surveys were blocks or groups of blocks; adjacent blocks on the listing were combined when necessary to ensure that each block had at least 20 housing units.

The list of census blocks was updated prior to the sample selection for the 1985-86 household survey. The purposes of the update were: (1) to reflect changes, since 1981, in administrative divisions and subdivisions, including the creation of one new district from parts of existing districts, (2) to reflect changes, including creation of new settlements and inundation of existing settlements, resulting from the progress of the Mahaweli development projects, and (3) to adjust housing unit counts or create new blocks in other areas of rapid growth, especially that resulting from the government's extensive programme for construction of new housing. Inputs for items (1) and (3) were compiled by the district statistical officers on the basis of local inquiries. Inputs for item (2) were obtained by statistical officers of the central office through contacts with project officials and field-work as needed. All changes were entered manually on the census block listing. As a result of these efforts, some 163 new "units" of about 175 families each were added to the frame in the Mahaweli settlement areas in five districts and the existence of 34,550 new housing units in 24 of the 25 districts was reflected by creating new blocks or increasing the measures of size for existing blocks.

<u>Master sample</u>: The sample PSUs (census blocks or groups of blocks) for the 1984-85 and the 1985-86 surveys were selected independently. However, it is planned to use the sample PSUs selected for the 1985-86 survey in the 1986-87 survey as well; therefore, this sample can be considered a master sample.

The stratification for the master sample was by district and by sector within district. In the urban sector of the Colombo District, four separate strata were established: for Colombo, Dehiwela/Mt. Lavinia, Kotte and the remainder of the urban sector. PSUs were allocated to districts in proportion to the square root of 1981 population. The urban strata were over-sampled by allocating approximately one-third of the sample PSUs to these strata, which have about one-fifth of the total population. The allocation of PSUs to districts and to the urban strata in the Colombo District was made in multiples of 12, to facilitate the assignment of PSUs to the 12 monthly rounds of the survey. Within each stratum, the assigned number of PSUs was selected with probability proportionate to size (using the census or adjusted housing unit counts), with replacement. The selected PSUs were assigned to monthly rounds by a systematic random procedure.

<u>Subsampling from the master sample</u>: For the 1985-86 survey, a listing of housing units was prepared for each PSU about two months prior to the scheduled interviewing. The housing units were ordered by number of persons and a systematic random sample of 10 housing units was selected. Interviews were conducted for all households in the sample housing units.

For the 1986-87 survey, it is expected that the housing unit listings for the sample blocks will be updated prior to sample selection. <u>Remarks</u>: The main benefit gained from the use of the master sample is that only an update of listings for sample PSUs will be required for the 1986-87 survey, rather than completely fresh listings. The update task is made easier by the fact that housing units, rather than households, were used as the listing units (and USUs). The use of the same PSUs for both surveys will also improve the reliability of estimates of change for items common to the two surveys.

References:

- Rao, M.V.S.
 1983 Report on Mission to Sri Lanka (1 to 16 December 1983) ESCAP Statistics Division.
- Jabine, T.B.
 1984 Report on Mission to Sri Lanka, 7 June to 6 July 1984, UN Statistical Office.
- 3. Rao, M.V.S. 1984 Report on Mission to Sri Lanka (26 November to 14 December 1984), ESCAP Statistics Division.
- Jabine, T.B. 1985 Report on Mission to Sri Lanka, 21 May to 14 June, 1985, UN Statistical Office.

CASE-STUDY: Thailand

<u>Type of case-study</u>: This case-study describes the design of a master sample for use in conducting multiple surveys and survey rounds during a single year.

Basis for case-study: This case-study is based primarily on the reports of a technical adviser who undertook three missions to Thailand during 1981-1983. The adviser assisted the National Statistical Office (NSO) of Thailand with the development of an IHSP under an NHSCP project. At the start of the period covered by the NHSCP project, the NSO had already had several years of experience conducting a labour force survey with semi-annual rounds and several <u>ad hoc</u> household surveys on topics such as population change, migration, health, education, income and expenditures.

These household surveys were already integrated to some extent by sharing field staff and central facilities. Household listings prepared annually for the labour force surveys were also used as a frame for some of the other surveys. A major goal of the NHSCP project was to further integrate the NSO's programme of household surveys. A plan was developed to undertake a periodic multi-subject survey, to be known as the Multipurpose Survey (MS). This survey would include the basic demographic and social items and a standard labour force module in each round. Other compatible topics would be added in selected rounds.

This case-study is based on the proposed design of the MS as of mid-1983; it does not reflect any subsequent changes that may have occurred. The MS was scheduled to begin in calendar 1984, however, not all features of the proposed design could be implemented at that time.

<u>Substantive requirements</u>: The survey population for the MS was to be the non-institutional population of Thailand, excluding a small population in hill tribes who are not governed under the normal administrative structure. Data on core demographic and social variables and labour force status (based primarily on a one-week reference period) were to be collected in quarterly rounds. A supplemental module on television viewing, radio listening and newspaper readership was to be included in one of the four MS rounds in 1984. For subsequent years, supplemental modules were planned on topics such as health, education and use of leisure time.

It was intended that the MS sample should be designed to produce separate estimates for the Bangkok Metropolis and the four regions of Thailand: Central, North, Northeast and South. For each region, separate estimates were desired for municipal areas, sanitary districts (smaller urban population clusters, see reference 4) and villages. Operating environment and constraints: The Kingdom of Thailand had an estimated population of 51,724,000 as of July 1984 and a population density of 100 persons per square kilometre. The country is divided into 72 provinces plus the Bangkok Metropolis. Each province is divided into districts and subdistricts. Except in municipal areas, the subdistricts are divided into villages. Some population concentrations outside of municipal areas are designated as sanitary districts; these may consist of anything from part of a village to an entire district.

A census of population was conducted in 1980, and the municipal area maps and small area population and household counts from the census were available for use in developing a frame. In addition, the NSO conducts an annual village survey in which village population counts and other data are obtained from village headmen.

The NSO has a field office in each province, with a permanent staff consisting of a supervisor and 3 to 10 full-time interviewers, depending on the population of the province. There are also 150 full-time interviewers in the Bangkok Metropolis. The interviewers are responsible for field listing and interviewing for a variety of demographic and economic surveys.

<u>Master sampling frame</u>: Separate master sampling frames are maintained for the municipal areas and the remainder of the country.

The municipal area frame is based on the 1980 census. For the census, each municipal area had been divided into enumeration districts (EDs) and blocks. The blocks, which are subdivisions of EDs, are used to form PSUs for most household surveys. Detailed master maps are available, showing the boundaries and identification numbers of EDs and blocks. Preliminary population counts by ED and block were available shortly after the 1980 census.

The master sampling frame for the remainder of the country, outside of municipal areas, is a list of villages which has been maintained by the NSO since the early 1960s and is periodically updated, primarily on the basis of information obtained from the Department of Local Administration, Ministry of Interior. The numbers of districts, subdistricts, and villages have been increasing at a fairly rapid pace, roughly two percent per year, over the past several years (reference 4). Typically, although not always, new villages are formed by splitting existing villages.

In general, villages do not have clearly defined boundaries and detailed maps are not available. Prior to the 1980 census, maps were prepared for the then-existing sanitary districts and for large villages (300 or more dwelling units) not in sanitary districts.

The village lists are ordered by province, district and subdistrict. Villages entirely or partly in sanitary districts are

identified. For villages existing at the time of the 1980 census, the census population and household counts are available. Later counts are available from the annual village surveys, subject to some non-response.

<u>Master sample</u>: The proposed annual master sample was to be a sample of about 4,230 PSUs: blocks in municipal areas and villages elsewhere. Primary stratification was based on regions (groups of provinces), as follows: Bangkok Metropolis, remainder of Greater Bangkok (three provinces in the Central Region), remainder of Central Region, North Region, Northeast Region and South Region. Secondary strata within each of the primary strata consisted of municipal areas, villages entirely or partly in sanitary districts, and other villages.

Allocation of sample PSUs to primary strata would be roughly in proportion to population. Within each of the primary strata, the municipal areas and the villages in sanitary districts would be oversampled in relation to other villages in order to provide sufficiently reliable estimates for each of the secondary strata.

In the municipal area strata, blocks below a specified size would be combined with adjacent blocks to form PSUs prior to selection. In the sanitary district and village strata this would not be done; however, small villages, e.g., those with fewer than 100 persons, would be given a minimum measure of size, say 100, to protect against errors in frame counts.

In each secondary stratum, the allotted number of PSUs would be selected systematically, with probability proportionate to size. For the municipal area stratum, census block counts would be used as measures of size, subject to updating in areas with large amounts of new construction. For the other strata, the latest available counts from the annual village survey would be used.

A procedure would be established for subdividing and subsampling large sample PSUs prior to listing. Existing maps would be used for this purpose when available; in other cases (mostly for villages outside of sanitary districts) staff of the mapping unit or the field offices would prepare the necessary sketches.

<u>Subsampling from the master sample</u>: A proposal for consideration was that the sample of PSUs (blocks and villages) for the MS be divided systematically into five random subgroups and that two of these five subgroups be included in the sample for each of the four rounds, with 50 percent overlap between rounds, as follows:

Random subgroups

Survey Round	Listed and sampled prior to round	Sample interviewed in round
1	1,2	1,2
2	3	2,3
3	4	3,4
4	5	4,5

The sample for each round would be roughly half the size of that used in the semi-annual rounds of the labour force survey in prior years.

For each PSU to be included in a survey round for the first time, a listing of housing units (including vacant residential units) would be prepared shortly before the scheduled interview period. A fixed number of housing units, 15 in the municipal-area and sanitary-district PSUs and 10 in the other PSUs, would be selected from the listings and all households associated with those housing units would be interviewed. For PSUs remaining in the sample for a second round, the same sample of housing units would be used in the second round.

To accommodate the sample requirements for a special migration survey conducted annually in the Bangkok Metropolis, it was proposed that listings in that area prior to the first round include screening questions to identify housing units with one or more migrants (as defined for the survey). All eligible housing units identified in this way would be included in the sample for the migration survey.

An alternate proposal was to prepare listings for all sample PSUs prior to the first quarterly round and to include a sample of housing units or households from every PSU in each round. To keep the interviewing work-load at the desired level, only one-half as many housing units or households per round would be selected in each PSU, i.e., 7 or 8 in municipal-area and sanitary-district PSUs and 5 in other PSUs. This design could be used with or without overlap between rounds in the housing unit or household samples.

Compared with the first proposal, the second would have produced quarterly estimates with smaller sampling errors, because of the larger number of PSUs included in each quarter's sample. A second advantage would have been the availability of a larger sample for the annual migration survey in the Bangkok Metropolis. However, the second approach also had three disadvantages: uneven distribution, over the year, of the work-load for listing PSUs; higher travel costs (because of the larger number of PSUs to be visited in each quarterly round); and a longer interval, on the average, between listing and interviewing of housing units or households selected from the listings. <u>Remarks</u>: Several aspects of the proposed MS design and its relation to sample designs used by NSO for household surveys up to that time are of interest:

- 1. Earlier household surveys used by NSO used a three-stage design with districts (<u>amphoes</u>) as PSUs, blocks and villages as SSUs, and households as USUs. After the permanent field staff was established and the conditions of local travel improved in the 1960s and 1970s, it became evident that a two-stage design would be more efficient for the larger surveys; hence, a two-stage design, with blocks and villages as PSUs, was adopted for the labour force survey in the early 1980s. Households had been used as USUs in all surveys through 1983. The switch to housing units as USUs was proposed in part to reduce the cost of listing and in part because experience in using a sample of households for survey rounds six months apart suggested that housing units, being more stable, would work better for samples to be used in more than one round of a survey.
- 2. Through 1984, in the two-stage designs using blocks and villages as PSUs, the PSUs had been selected with equal probability. With the continuing creation of new villages, it is somewhat difficult to obtain reasonably up-to-date measures of size for every PSU, hence equal probability selection was used primarily as a matter of convenience. However, this design was clearly less efficient than one using selection of first-stage units with probability proportionate to size, and it was considered worthwhile to invest the effort needed to obtain usable measures of size from the annual village survey, the 1980 census, and other sources.
- 3. An initial sample design proposal for the MS called for the selection of a two-stage self-weighting sample within each of the secondary strata. However, experience in a test of the MS design and procedures in two provinces in January 1983 indicated that this complicated the field sampling procedures and made it difficult to control interviewer work-loads. Since special PSU weights would have been needed for other reasons anyway (to deal with subsampling of large PSUs and adjustments for non-response), it was decided to sample a predetermined number of housing units from each sample PSU.
- 4. The potential existed for further efficiencies (reduced listing costs, better estimates of change) by designing a master sample to be used for a period longer than one year. However, this would have required a relatively complex design in order to deal with the continuing creation of new villages; so the one-year approach was preferred for the time being.

References:

- Jabine, T.
 1981 A report on plans to develop an integrated system of household surveys for Thailand. American Public Health Association, Washington, D.C.
- 2. Jabine, T.
 - 1982 A progress report on the development of an integrated system of household surveys for Thailand. American Public Health Association, Washington, D.C.
- 3. Jabine, T.
 - 1983 The development of an integrated system of household surveys for Thailand: mid-term project review. American Public Health Association, Washington, D.C.
- 4. Skunasingha, A. and Jabine, T.
 - 1983 The administrative subdivisions of Thailand and their relation to household surveys. International Statistical Institute, 44th Session, <u>Contributed Papers</u>, Vol. 2: 664-667.

CASE-STUDY: United States of America

Type of case-study: This case-study describes the U.S. master sample of agriculture. Although designed primarily for use in surveys of agricultural holdings (farms), the master sample of agriculture was also used for household surveys. It is described here because of its historical importance and interest as the first application of the master sample concept.

Basis for case-study: 1945 articles by Jessen and King (References 2 and 3) provide details of the origin and initial development of the master sample. A 1984 article by Fuller (Reference 1) summarizes this information and describes subsequent uses of the master sample.

<u>Substantive requirements</u>: The idea of developing a master sample is credited to Rensis Likert, who was then employed by the Bureau of Agricultural Economics (BAE) in the U.S. Department of Agriculture. The BAE had been conducting a variety of surveys of farms and farm population and, according to King (Reference 2, p. 43):

It was evident that there was need for a procedure which would provide effective samples for various studies and more particularly which would provide the accumulation of data relating to a representative group of farms. By drawing subsamples for different studies from a large Master Sample and systematically accumulating the data for these farms and farm families, many important interrelationships affecting farm production, income and living which cannot now be studied economically could be analyzed. Thus a Master Sample could become a device for integrating and improving the effectiveness of the research of the Bureau.

The initial plan was to design a set of areas which would provide a national sample of about 5,000 farms. However, the target sample size was increased, first to 25,000 and then to a sample of 300,000 farms, about five percent of all U.S. farms at that time, in order to meet requirements for use of the sample in the 1945 census of agriculture. The U.S. Bureau of the Census, having been informed of the planned master sample, wanted to use it as a method of obtaining a sample of farms that would be asked to respond to a supplemental questionnaire that was to be part of the census.

The expansion of the master sample to meet the Census Bureau's requirements for the 1945 agriculture census enhanced the possibility of its use for many purposes. Samples in most states were large enough to provide reliable data for farms and farm population at the state level, and the master sample was used in that way in several states. An expanded version of the master sample, covering all population outside of "block cities" (larger cities for which block statistics were published in the population census), was used to select samples for the

survey of the labour force that is now known as the Current Population Survey. In the early years after the development of the master sample, the frame and sample materials were heavily used; the Department of Agriculture used them to draw 60 to 80 samples annually in the first ten years. Even today, updated versions of the original master sample frame are used extensively for surveys of farms and rural population.

Operating environment and constraints: In 1940, the United States of America had a resident population of 131,669,000 with a population density of 14.4 per square kilometre. The population in places with less than 2,500 population and other rural territory was 57,246,000. There were about 6,100,000 farms and 30,547,000 persons living on farms.

The United States at that time consisted of 48 states and the District of Columbia, plus territories and possessions. The 48 states were subdivided into 3,070 counties.

During the 1930s and early 1940s, in connection with public works programmes, a large-scale county highway map had been prepared for almost every county. In addition to showing roads, railroads, rivers, and boundaries of political subdivisions, the county highway maps showed the location of dwellings and other buildings in rural areas. These maps were the basic materials used in developing the frame for the master sample. Also available, when work on the master sample started, were aerial photographs covering about 90 percent of the agricultural area of the United States. These photographs were used in some areas to establish boundaries of the USUs (area segments, see below) and also proved useful in applications of the master sample that required identification and measurement of individual fields.

Master sampling frame: Using the county highway maps, each county was subdivided into three strata: incorporated areas, densely populated unincorporated areas and sparsely populated unincorporated areas, generally referred to as "open country". About 91 percent of all farms were located in the open-country stratum. The following details refer to the open-country stratum; for details on the treatment of the other strata, see reference 2.

In each county the open-country portion was divided into "count units" containing from 6 to 30 farms each. The count units were established within minor civil divisions, using natural boundaries such as roads, rivers and streams. Each minor civil division (the next political subdivision below the county level, usually called a town or township) with open-country area was assigned at least one count unit, even if it appeared to have fewer than six farms.

The count units were listed by minor civil division, and for each count unit the following counts were recorded: indicated number of farms, indicated number of dwellings and number of sample units assigned to the count unit. The number of sampling units was assigned on the basis of rules intended to result in sampling units averaging between four and eight farms, depending on the region of the country and on whether aerial photographs were available.

The resulting basic materials for the master sampling frame consisted of the count-unit listings plus the county highway maps showing the boundaries and identification of the count units.

<u>Master sample</u>: The master sample for the open-country stratum was selected in two stages. In the first stage, count units were selected systematically, with probability proportionate to the number of sampling units assigned to them. A sampling interval of 18 was used. In preparation for the second stage of selection, all selected count units with two or more assigned sampling units were subdivided, using natural boundaries, into the assigned number of sampling units. One of these sampling units was then selected at random from each selected count unit.

The final sample for the open-country stratum consisted of about 60,400 sampling units, usually referred to as sample areas or area segments. The average area of these segments ranged from less than two square kilometres in the State of Indiana to about 280 square kilometres in Nevada. For the other two strata, a total sample of about 6,600 area segments, designed to include 1 in 18 of the farms in these strata, was selected.

Subsampling from the master sample: The only application of the full master sample was its use to collect supplemental data for a sample of farms in the 1945 Census of Agriculture. All other uses involved subsampling. Detailed information on designs for individual applications is not readily available. Some national surveys used counties as PSU's. Within the sample PSU's, it would be possible to use all or a subset of the master sample area segments or, if a larger sample were needed, to use the master sample frame materials to select additional segments. For surveys at the state level, it was probably convenient in most instances to select a one-stage subsample of master sample area segments.

<u>Remarks</u>: By any standard, the development and use of the master sample of agriculture must be judged a successful undertaking. It pioneered the use of area sampling with an attempt to optimize the sampling units in relation to travel time and interviewing costs. Substantial resources were devoted to frame construction and sample selection, but it was possible to use the resulting materials for a large number of surveys. The area sample was versatile: with appropriate rules of association it could be used to select samples of farms, fields, dwelling units or households. For most of the open-country segments, the availability of aerial photographs made it possible to measure the area of fields, using planimetering techniques.

The two-stage design used to select the master sample substantially reduced the amount of preparatory work needed, since only the selected count units had to be subdivided into sampling units. References:

- Fuller, W.
 1984 The master sample of agriculture. In <u>Statistics: An</u> <u>Appraisal</u>, David, H.A. and David, H.T., eds. Ames: Iowa State University Press.
- Jessen, R. 1945 The master sample of agriculture, II. Design. <u>Journal of the</u> <u>American Statistical Association</u>, 40 (229): 46-56.
- 3. King, A.
 - 1945 The master sample of agriculture, I. Development and use. Journal of the American Statistical Association, 40 (229); 38-45.

ANNEX II

BIBLIOGRAPHY

Part 1 - Recommended additional reading

The following books, manuals and articles are recommended to readers who would like to have more information about the topics covered in this technical study. The items in this section are organized by topic.

A. Integrated household survey programmes

Two papers presented at the 1983 session of the International Statistical Institute provide an excellent introduction to the concept of integration of household surveys and to the application of that concept in the National Household Survey Capability Programme (NHSCP):

- 1. Foreman, E.K. (1983). Integrated programmes of household surveys: design aspects. <u>Bulletin of the International</u> <u>Statistical Institute</u>, 50(2): 1344-1362.
- 2. United Nations Statistical Office (1983). National Household Survey Capability Programme: selected issues of design and implementation. <u>Bulletin of the International Statistical</u> Institute, 50(2): 1363-1380.

Four technical studies prepared specifically for the NHSCP have dealt with key aspects of household survey design, methodology and infrastructure. They are:

- 3. United Nations (1982). Survey Data Processing: A Review of Issues and Procedures, DP/UN/INT-81-041/1.
- 4. United Nations (1982). <u>Non-sampling Errors in Household</u> Surveys: Sources, Assessment and Control, DP/UN/INT-81-041/2.
- 5. United Nations (1983). Role of the NHSCP in Providing Health Information in Developing Countries, DP/UN/INT/81-041/3.
- 6. United Nations (1985). <u>Development and Design of Survey</u> Questionnaires, INT-84-014.

Three relevant papers on "Problems of development of survey infrastructure" were presented at the 1985 session of the International Statistical Institute:

7. de Graft-Johnson, K. (1985). Issues and problems encountered in building up survey infrastructure. Paper 17.1.

- 8. Coker, J. and Jones, G. (1985). Field organizations: tasks and practices. Paper 17.3.
- 9. Carlson, B. and Sadowsky, G. (1985). Survey data processing infrastructure in developing countries: selected problems and issues. Paper 17.4.

Two technical papers on the United States Current Population Survey (CPS) provide the most detailed readily available description of an IHSP design used in a developed country. The CPS uses design A as defined in Chapter III. section D of this study, i.e., a single, periodic multi-subject survey.

- U.S. Bureau of the Census (1963). <u>The Current Population</u> <u>Survey, a Report on Methodology</u>. <u>Technical Paper 7</u>. Washington, D.C.: U.S. Department of Commerce.
- U.S. Bureau of the Census (1978). <u>The Current Population</u> <u>Survey: Design and Methodology</u>. <u>Technical Paper 40</u>. Washington, D.C.: U.S. Department of Commerce.

The U.S. Census Bureau also prepared, for training purposes, a detailed description of a household survey programme designed for an imaginary developing country called Atlantida. The design used for this case-study is also design A, a single periodic multi-subject survey.

12. U.S. Bureau of the Census (1966). <u>Atlantida: A Case Study in Household Sample Surveys</u>. Series ISPO-1. Washington, D.C.: U.S. Department of Commerce. The series has 14 separate units. Of most interest in connection with this technical study are: Unit I. <u>Survey objectives and description of country</u>. Unit II. <u>Content and design of household surveys</u>. Unit IV <u>Sample design</u>

B. Sampling frames for household surveys

The following articles and reports deal with various aspects of sampling frames for household surveys:

- 1. Dalenius, T. (1983). Frame Construction Techniques for Sample Surveys. Swedish Agency for Research Cooperation with Developing Countries. A broad treatment of the subject, with emphasis on rules of association. Examples cover both household and business surveys.
- Hansen, M., Hurwitz, W. and Jabine, T. (1963). The use of imperfect lists for probability sampling at the U.S. Bureau of the Census. <u>Bulletin of the International Statistical</u> <u>Institute</u>, 40(1): 497-516. This paper identifies problems in

using imperfect lists as sampling frames and gives suggestions and examples of how the problems have been overcome in practice.

- 3. Hartley, H. (1974). Multiple frame methodology and selected applications. <u>Sankhya</u>, 36, Series C, Part 3: 99-118. In this paper, Hartley summarizes the concepts and results of multiple frame methodology presented in earlier papers by himself and other authors.
- 4. Lessler, J. (1980). Errors associated with the frame. <u>American Statistical Association Proceedings, Section on Survey</u> <u>Research Methods</u>, 125-130. This paper presents a detailed and informative classification of frame errors and discusses their effects.
- 5. Monroe, J. and Finkner, A. (1959). <u>Handbook of Area Sampling</u>. Philadelphia: Chilton. This manual provides a practical description of the uses of census data and maps to establish area frames for sample surveys.
- 6. Szameitat, K. and Schaffer, K. (1963). Imperfect frames in statistics and the consequences for their use in sampling. <u>Bulletin of the International Statistical Institute</u>, 40 (1): 497-517. The authors present a mathematical model for analysis of the errors associated with frame imperfections.

The following two manuals on mapping for censuses and surveys should be useful to persons interested in the development of area sample frames for use in household surveys:

- Cooke, D. (1971). <u>Mapping and House Numbering</u>, Laboratory for Population Statistics Manual Series, No. 1. University of North Carolina.
- 8. U.S. Bureau of the Census (1978). <u>Mapping for Censuses and</u> <u>Surveys</u>, Statistical Training Document ISP-TR-3. Washington, D.C.: U.S. Department of Commerce.
- C. Master samples

In the literature on surveys there are only a few items that deal explicitly with master samples. Three of these items are about the United States master sample of agriculture:

- Fuller, W. (1984). The master sample of agriculture. In <u>Statistics: An Appraisal</u>, David, H.A. and David, H.T., eds. Ames: Iowa State University Press.
- Jessen, R. (1945). The master sample of agriculture, II. Design. <u>Journal of the American Statistical Association</u>, 40(229): 46-56.

3. King, A. (1945). The master sample of agriculture, I. Development and use. Journal of the American Statistical Association, 40(229):38-45.

Other items are:

- 4. Deville, J. and Roy, G. (1984). L'echantillon-maitre fait peau neuve. <u>Courrier des Statistiques</u>, 29: 24-26. This article describes the design of a master sample selected for use in household surveys (excluding the labour force survey) during the ten-year period following France's 1982 census of population.
- 5. Murthy, M. (1981). Master sampling frame and master sample for household sample surveys in developing countries. <u>Rural</u> <u>Demography</u>, 8(1): 13-27. This article presents, in a compact form, many of the basic concepts discussed in this technical study. Procedures for subsampling from a master sample are illustrated, using a master sample of villages in a subdivision of Bangladesh.

D. Household survey methods

There are many texts and manuals on sampling. Those listed here include some discussion of master samples or related topics, such as sampling for time series. Exclusion of a publication from this list does not carry any implications about its relative merits as a general text or manual on sampling.

- Cochran, W. (1963). <u>Sampling Techniques</u> (2nd ed.). New York: J. Wiley.
- Hansen, M., Hurwitz, W. and Madow, W. (1953). <u>Sample Survey</u> <u>Theory and Methods</u>, Volume I, Methods and Applications. New York: J. Wiley
- 3. Kish, L. (1965). Survey Sampling. New York: J. Wiley.
- 4. Yates, F. (1949). <u>Sampling Methods for Censuses and Surveys</u>. London: Charles Griffin.

There are also many texts and manuals that provide a broad treatment of survey design and methodology, including sampling. The three that follow are included because they are intended to apply primarily to surveys in developing countries:

5. Casley, D. and Lury, D. (1981). <u>Data Collection in Developing</u> <u>Countries</u>. Oxford: Oxford University Press.

- 6. United Nations (1984). <u>Handbook of Household Surveys</u> (Revised Edition). Studies in Methods F, No. 31 (ST/ESA/STAT/SER. F/31).
- World Fertility Survey (1975-1977) <u>Basic Documentation</u>. Consists of 12 manuals covering various aspects of survey design and operations. Of particular relevance to this technical study are No. 2, <u>Survey Organization Manual</u> and No. 3, <u>Manual on Sample Design</u>. Both are available in English, French, Spanish and Arabic.

Part 2 - References

This section of Annex II lists all references appearing in the text of this technical study. References used to document the case-studies in Annex I are included only if they are readily available in published form.

Bailar, B.

1975. The effects of rotation group bias on estimates from panel surveys. <u>Journal of the American Statistical Association</u>, 70:23-30.

Cooke, D.

- 1971. <u>Mapping and House Numbering</u>, Laboratory for Population Statistics Manual Series No. 1. University of North Carolina.
- Duncan, J. and Shelton, W.

1978. <u>Revolution in United States Government Statistics</u>, <u>1926-1976</u>. Office of Federal Statistical Policy and Standards, U.S. Department of Commerce.

- Fellegi, I. and Sunter A 1974 Balance between different sources of survey errors - some Canadian experiences. Sankhya, C.: 119-142.
- Fuller, W.
- 1984. The master sample of agriculture. In <u>Statistics: An</u> <u>Appraisal</u>, David, H.A. and David, H.T., eds. Ames: Iowa State University Press.
- Garrett, J. and Stanecki, K.
 - 1983. A master sample approach to the Moroccan intercensal survey program. <u>American Statistical Association</u> <u>Proceedings of the Social Statistics Section</u>.
- Hansen, M., Hurwitz, W. and Madow, W. 1953. <u>Sample Survey Methods and Theory</u>, <u>Volume I, Methods and Applications</u>. New York: J. Wiley.
- Jessen, R.J.
- 1945. The master sample of agriculture, II. Design. <u>Journal of</u> the American Statistical Association, 40(229): 46-56.

Keyfitz, N.

1951. Sampling with probabilities proportionate to size: adjustments for changes in probabilities. <u>Journal of the</u> <u>American Statistical Association</u>, 46: 105-109.

King, A.J. 1945. The master sample of agriculture, I. Development and Journal of the American Statistical Association. use. 40(229): 38-45. Kish, L. 1965. Survey Sampling. New York: J. Wiley. Murthy, M.N. 1969. Population census as the source of 'sampling frame in India. Sankhya, B: 1-12. Neyman, J. On the two different aspects of the representative method: 1934. the method of stratified sampling and the method of purposive selection. Journal of the Royal Statistical Society, 97:558-625. Rao, V.R. and Sastry, N.S. 1975. Evolution of total survey design: the Indian experience. Bulletin of the International Statistical Institute, 46(1): 208-220. Rustemeyer, A. 1977. Measuring interviewer performance in mock interviews. American Statistical Association Proceedings, Social Statistics Section: 341-346. Sirken, M. 1970. Household surveys with multiplicity. Journal of the American Statistical Association, 65: 257-266. Sirken, M. 1975. Network surveys. Bulletin of the International Statistical Institute, XLVI (4): 332-342. Skunasingha, A. and Jabine, T.B. The administrative subdivisions of Thailand and their 1983. relation to household surveys. International Statistical Institute, 44th Session, Contributed Papers, 2: 664-667. United Nations 1980a Principles and Recommendations for Population and Housing Censuses. Statistical Papers, Series M, No. 67 (Sales No. 80.XVII.8). United Nations 1980b The National Household Survey Capability Programme: Prospectus (DP/UN/INT-79-020/1).
United Nations

1982a. <u>Non-sampling Errors in Household Surveys: Sources,</u> Assessment and Control (DP/UN/INT-81-041/2).

United Nations

1982b. <u>Survey Data Processing: A Review of Issues and Procedures</u> (DP/UN/INT-81-041/1).

United Nations

1984 <u>Handbook of Household Surveys</u> (Revised Edition). Studies in Methods Series F, No. 31 (ST/ESA/STAT/SER. F/31).

United Nations Statistical Office

1983. National Household Survey Capability Programme: selected issues of design and implementation. <u>Bulletin of the</u> <u>International Statistical Institute</u>, 50(2): 1363-1380.

U.S. Bureau of the Census

- 1977. <u>The Current Population Survey: Design and Methodology</u>, Technical Paper 40. Washington, D.C.: U.S. Department of Commerce.
- U.S. Bureau of the Census
 - 1978. <u>Mapping for Censuses and Surveys</u>, Statistical Training Document ISP-TR-3. Washington, D.C.: U.S. Department of Commerce.

World Fertility Survey

1975. <u>Manual on Sample Design</u>, Basic Documentation, No. 3. The Hague: International Statistical Institute. Chapter 5. Sampling Frames.

Yates, F.

1949. <u>Sampling Methods for Censuses and Surveys</u>. London: Charles Griffin.

Printed in U.S.A.

40737-July 1991-500