METHODOLOGICAL REPORTS 3

DHS Demographic and Health Surveys

Methods of Estimating Contraceptive Prevalence Rates for Small Areas

Applications in The Dominican Republic and Kenya

MACRO INTERNATIONAL INC

The Demographic and Health Surveys (DHS) is a 13-year project to assist government and private agencies in developing countries to conduct national sample surveys on population and maternal and child health. Funded primarily by the United States Agency for International Development (USAID), DHS is administered by Macro International Inc. in Columbia, Maryland.

The main objectives of the DHS program are (1) to promote widespread dissemination and utilization of DHS data among policymakers, (2) to expand the international population and health database, (3) to advance survey methodology, and (4) to develop in participating countries the skills and resources necessary to conduct high-quality demographic and health surveys.

For information about the Demographic and Health Surveys program, write to DHS, Macro International Inc., 11785 Beltsville Drive, Suite 300, Calverton, MD 20705, U.S.A. (Telephone 301-572-0200; Telefax 301-572-0999).

Demographic and Health Surveys Methodological Report No. 3

Methods of Estimating Contraceptive Prevalence Rates for Small Areas: Applications in The Dominican Republic and Kenya

ALFREDO ALIAGA
PRADIP K. MUHURI

Macro International Inc.
Calverton, Maryland USA

June 1994

Recommended citation:

Aliaga, Alfredo and Pradip K. Muhuri. 1994. *Methods of Estimating Contraceptive Prevalence Rates for Small Areas: Applications in The Dominican Republic and Kenya.* DHS Methodological Reports No. 3. Calverton, Maryland: Macro International Inc.

Contents

Figures

1 Introduction

Although direct estimates are generally used to determine contraceptive prevalence rates at the national level, such estimates are often unreliable for small areas because the samples are of insufficient size to be representative. Increasingly, however, program managers and policymakers in developing countries have expressed interest in obtaining estimates of basic demographic indicators for small areas in order to set targets, allocate resources, and monitor the performance of family planning and maternal and child health programs (Muhuri and Rutstein, 1994).

Conducting nationally representative DHS surveys in countries that include geographical areas containing small populations is possible, however, it may not be cost-effective. In addition, nonsampling errors may increase with increased sample size. One solution to these problems is to rely on indirect methods to generate estimates for small population areas. This paper examines the suitability of two methods for calculating indirect estimates—the *synthetic* estimation procedure and the *regression-based* procedure. Described herein is the estimating equation and the application of each method to the Dominican Republic and Kenya. These two countries were selected because DHS surveys had already been carried out and the samples available were relatively large (IEDP, ONAPLAN, and IRD, 1992; NCPD and IRD, 1989). It is possible, therefore, to compare *direct* estimates with the estimates derived from using the two indirect methods.

Estimates of contraceptive prevalence rates are provided for 30 provinces of the Dominican Republic and 32 districts of Kenya using the *synthetic* and *regression* methods. These are then compared with the results from the *direct* estimation procedure. It was found, after carrying out this comparison, that the *regression* approach is more suitable than the *synthetic* approach as an indirect method for estimating the prevalence of contraceptive use.

2 Definition of a Small Area

In the context of the DHS surveys, a *small area* is defined as a geographical administrative division that has a relatively small population and, therefore, has a corresponding sample size that is too small to provide accurate estimates of demographic parameters. In general, DHS provides estimates of basic demographic parameters, including contraceptive prevalence, that pertain to the nation as a whole, to urban and rural areas, and to the major regions of the country. Administrative divisions smaller than major regions are not usually adequately represented in the surveys.

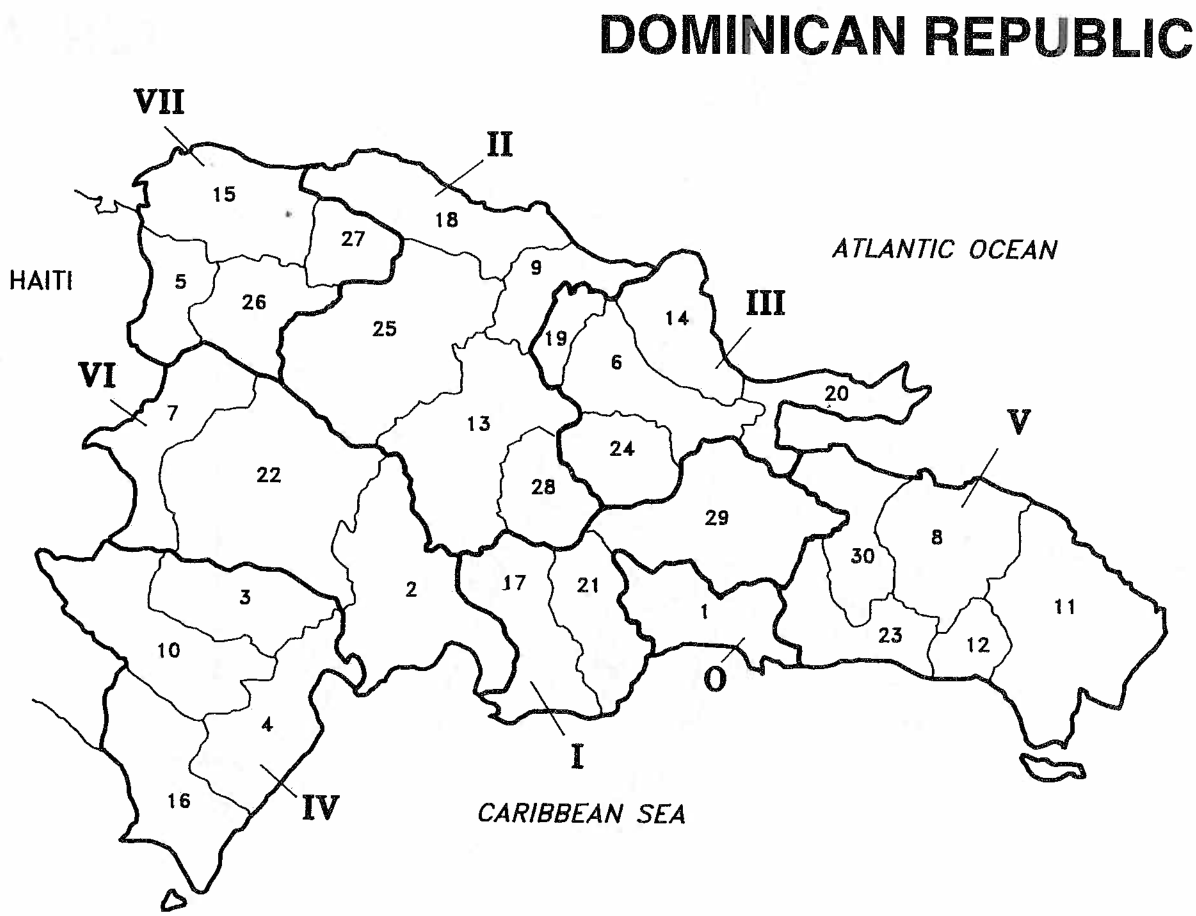
For the 1991 DHS survey, the Dominican Republic was divided into 8 regions consisting of 30 provinces (see Figure 2.1). The census enumeration area is the primary sampling unit (PSU). The sample included 396 PSUs; at least 40 PSUs were selected from each region. Kenya is divided into 8 provinces which contain 41 districts (see Figure 2.2). However, the 1989 Kenya DHS survey included only 34 districts in 7 provinces covering about 95 percent of the population. As in the Dominican Republic, the census enumeration area in Kenya is the PSU. From each of the seven provinces, at least 40 PSUs were selected. In both countries, the samples were selected in two stages. In the first stage, PSUs were systematically selected with probability proportional to size (the size equals the number of households in the PSU). In the second stage following an operation to list all households in each selected PSU, individual households were selected with the probability of selection inversely proportional to the PSU's size. This procedure is used in order to maintain a self-weighting sample within major sampling domains.[1] The sample of selected households in each PSU is called a *cluster*.

Provinces in the Dominican Republic and *districts* in Kenya are designated as small areas in this analysis. The goal of this exercise is to obtain estimates of the contraceptive prevalence rates for small areas, i.e., administrative geographical areas (districts/provinces) in which samples were not representative.

---

[1] For a detailed explanation of the procedure see the DHS *Sampling Manual* (Institute for Resource Development, 1987).

Figure 2.1 Map of the Dominican Republic showing regions and provinces, 1991



MAP KEY

0
1 DISTRITO NACIONAL

I
17 PERAVIA
21 SAN CRISTOBAL
29 MONTE PLATA

II
9 ESPAILLAT
13 LA VEGA
18 PUERTO PLATA
25 SANTIAGO
28 MONSEÑOR NOUEL

III
6 DUARTE
14 MARIA TRINIDAD SANCHEZ
19 SALCEDO
20 SAMANA
24 SANCHEZ RAMIREZ

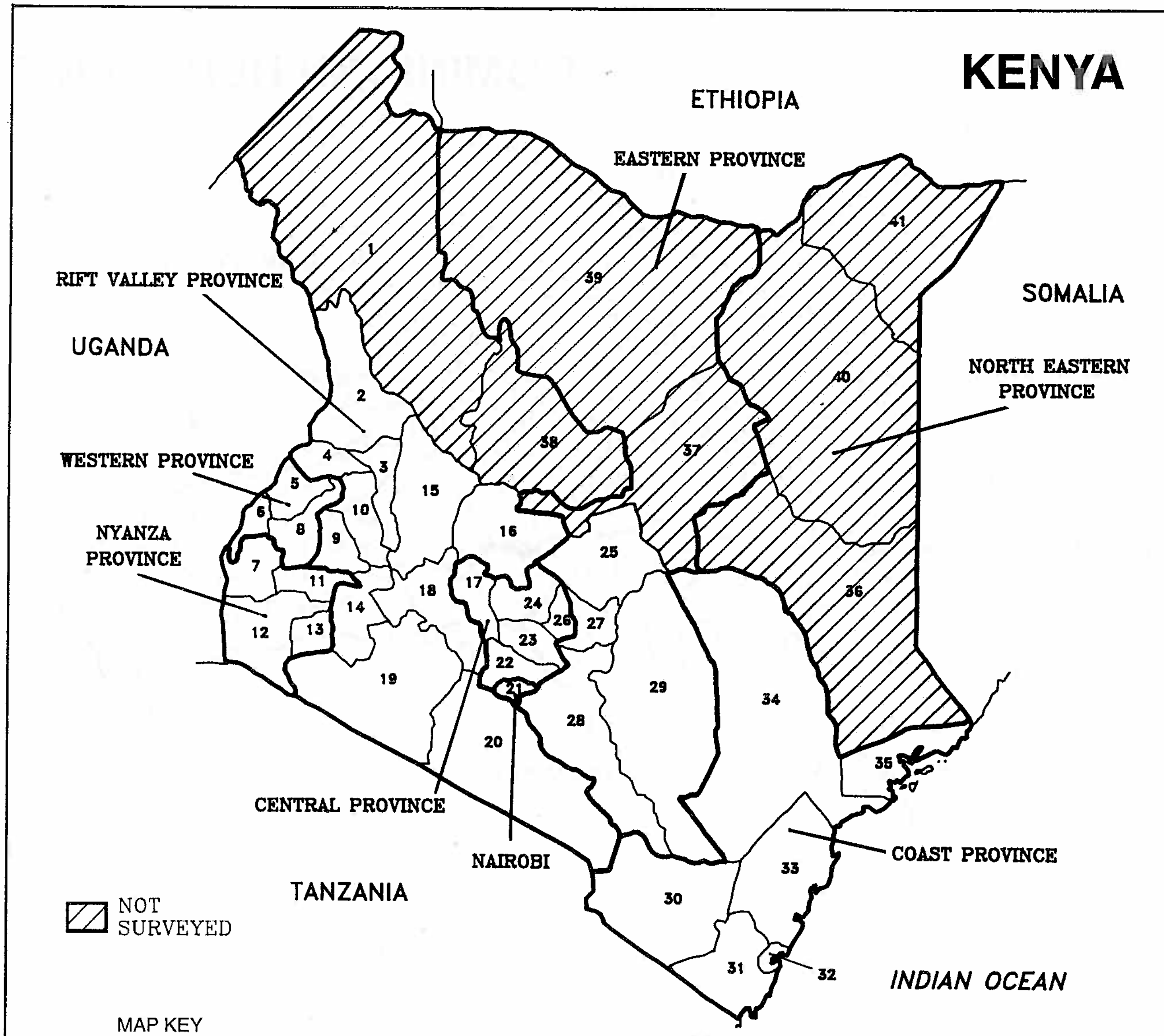
IV
3 BAHORUCO
4 BARAHONA
10 INDEPENDENCIA
16 PEDERNALES

V
8 EL SEIBO
11 LA ALTAGRACIA
12 LA ROMANA
23 SAN PEDRO DE MACORIS
30 HATO MAYOR

VI
2 AZUA
7 ELIAS PIÑA
22 SAN JUAN

VII
5 DAJABON
15 MONTECRISTI
26 SANTIAGO RODRIGUEZ
27 VALVERDE

Figure 2.2 Map of Kenya showing provinces and districts, 1989



MAP KEY

21 Nairobi

CENTRAL PROVINCE
22 Kiambu
26 Kirinyaga
23 Murang'a
17 Nyandarua
24 Nyeri

COAST PROVINCE
33 Kilifi
31 Kwale
35 Lamu
32 Mombasa
30 Taita
34 Tana River

EASTERN PROVINCE
27 Embu
37 Isiolo
29 Kitui
28 Machakos
39 Marsabit
25 Meru

NORTHEASTERN PROVINCE
36 Garissa
41 Mandera
40 Wajir

NYANZA PROVINCE
13 Kisii
11 Kisumu
7 Siaya
12 South Nyanza

RIFT VALLEY PROVINCE
15 Baringo
3 Elgeyo Marakwet
20 Kajiado
14 Kericho
16 Laikipia
18 Nakuru
9 Nandi
19 Narok
38 Samburu
4 Trans Nzola
1 Turkana
10 Uasin Gishu
2 West Pokot

WESTERN PROVINCE
5 Bungoma
6 Busia
8 Kakamega

3 Direct Estimation

The contraceptive prevalence rate is defined as the percentage of married women age 15-49 who are currently using any contraceptive method. It is the ratio of the weighted number of women who are using contraception in an area to the weighted number of married women age 15-49. The formula for *direct estimation* of the contraceptive prevalence rate, r, is as follows:

$$r = \frac{\sum_{i=1}^{n} w_i \, y_i}{\sum_{i=1}^{n} w_i \, x_i}$$

where:

- r = the contraceptive prevalence rate in a given administrative area;
- w_i = the sampling weight for the ith woman;
- y_i = whether the ith woman is using contraception currently (1 for yes and 0 for no);
- x_i = the counting variable whose value is 1 for every currently married woman;
- n_i = the number of currently married women in the given administrative area.

The procedure is simple, but suitable *only* if the sample size is large. Using this *direct* procedure, the contraceptive prevalence rate is estimated for 7 regions and 30 provinces of the Dominican Republic and 8 provinces and 32 districts of Kenya. The sampling error, which is the square root of the sampling variance, and the relative error, which is the ratio of the sampling error to the estimate[2] (converted into percentages), are computed for regions and provinces in the Dominican Republic and for provinces and districts in Kenya. Sampling errors and relative errors were not calculated for 12 provinces of the Dominican Republic and for 14 districts of Kenya. Each of these units had fewer than 9 primary sampling units (PSUs) and, consequently, insufficient degrees of freedom for calculation of the sampling error.

The consistency of the direct estimates was tested by repeatedly calculating prevalence rates based on a varying number of PSUs within a province/district.

In the Dominican Republic, all provinces other than Distrito Nacional have fewer than 30 PSUs. The regional estimates, if they are based on more than 40 PSUs, seem to be reasonable (the larger the number of PSUs in the area, the better the estimate), with relative errors less than 7 percent; the only exception is Region VI for which the relative error of the estimate is 14 percent (see Table 3.1). The estimate for Distrito Nacional, with 65 PSUs, which is the only province under Region 0, is fairly consistent. Estimates for the four provinces which are based on 20-30 PSUs are in the critical region, meaning that they are close to being reasonable but do not have enough consistency. The relative error of those provinces ranges from 5 to 24 percent.

For Kenya, the province- and district-level estimates and other statistics are presented in Table 3.2. As in the Dominican Republic, the province-level prevalence estimates are consistent (the larger the number of PSUs in the area, the better the estimate), if they are based on more than 40 PSUs. Their relative errors range from 5 to 10 percent except for the Western province (15.9). There are 13 districts with 20 to 30 PSUs. The relative errors range from 4 to 23 percent (see Table 3.2).

From the analysis thus far, it was concluded that contraceptive prevalence rates for small areas that have more than 40 PSUs, estimated using the *direct* approach, yield fairly consistent rates.

[2] Sampling errors were calculated using CLUSTERS software.

Table 3.1 Direct estimates of contraceptive prevalence for married women: Dominican Republic, 1991

Region and province	Direct Estimate (r)	Sampling error (SE)	Relative error (SE/r)	Number of PSUs (m)	Number of cases	
					Unweighted (N)	Weighted (NW)
Region 0	**60.7**	**2.0**	**3.3**	**65**	**688**	**1450**
Distrito Nacional	60.7	2.0	3.3	65	688	1450
Region I	**50.8**	**2.9**	**5.8**	**48**	**471**	**250**
Peravia	48.7	5.8	11.9	14	134	86
San Cristóbal	55.9	3.7	6.6	23	243	112
Monte Plata	43.1	3.9	9.0	11	94	52
Region II	**61.0**	**2.1**	**3.5**	**63**	**643**	**749**
Santiago	62.1	3.5	5.7	29	289	327
Puerto Plata	62.8	5.3	8.5	10	99	102
La Vega	59.4	4.4	7.3	10	108	180
Espaillat	56.2	a	a	7	70	58
Monseñor Nouel	61.7	a	a	7	77	82
Region III	**57.4**	**3.7**	**6.5**	**50**	**567**	**392**
Salcedo	53.4	a	a	7	77	45
Duarte	57.2	4.5	7.9	19	220	101
María T. Sánchez	65.9	7.5	11.4	10	105	70
Samaná	44.6	a	a	5	53	21
Sanchez Ramirez	56.7	8.2	14.5	9	112	155
Region IV	**47.1**	**2.7**	**5.8**	**41**	**430**	**174**
Barahona	51.9	2.7	5.1	21	226	88
Pedernales	48.3	a	a	3	32	15
Bahoruco	38.8	6.5	16.8	11	104	48
Independencia	44.8	a	a	6	68	23
Region V	**50.6**	**3.4**	**6.7**	**45**	**524**	**730**
La Romana	44.0	7.9	17.9	12	171	241
La Altagracia	57.5	a	a	7	86	106
El Seybo	46.5	a	a	7	79	116
San P. de Macorís	54.2	3.5	6.4	13	135	174
Hato Mayor	58.0	a	a	6	53	93
Region VI	**39.7**	**5.7**	**14.4**	**43**	**491**	**193**
San Juan	34.1	8.2	24.0	22	236	106
Azua	52.3	4.5	8.6	15	183	62
Elías Piña	32.1	a	a	6	72	25
Region VII	**58.6**	**2.8**	**4.8**	**41**	**412**	**146**
Valverde	53.9	2.4	4.5	15	143	57
Santiago Rodriguez	73.6	a	a	7	63	22
Dajabón	51.8	a	a	7	78	30
Monte Cristi	62.8	4.4	7.1	12	128	37
TOTAL	56.4	1.2	2.1	396	4226	4083

[a] Sampling errors not calculated because fewer than 9 PSUs.

Table 3.2 Direct estimates of contraceptive prevalence for married women: Kenya, 1989

Province and district	Direct Estimate (r)	Sampling error (SE)	Relative error (SE/r)	Number of PSUs (m)	Number of cases Unweighted (N)	Number of cases Weighted (NW)
NAIROBI	**33.5**	**2.1**	**7.6**	**40**	**519**	**335**
CENTRAL	**39.5**	**1.7**	**5.7**	**84**	**787**	**648**
Kiambu	37.3	a	a	5	61	134
Kirinyaga	54.2	3.2	7.2	25	236	88
Murang'a	32.1	5.4	16.7	25	217	162
Nyandarua	39.2	a	a	4	30	83
Nyeri	40.8	3.2	8.7	25	243	181
COAST	**18.1**	**1.7**	**8.9**	**47**	**529**	**350**
Kilifi	10.8	1.7	19.4	24	330	125
Kwale	16.2	a	a	3	34	91
Mombasa	24.5	3.6	17.3	16	147	95
Taita	30.7	a	a	4	18	39
EASTERN	**40.2**	**2.1**	**7.3**	**52**	**561**	**804**
Embu	47.2	a	a	3	27	79
Kitui	41.3	a	a	3	59	186
Machakos	40.4	2.9	7.9	24	282	337
Meru	36.3	3.5	11.6	22	193	202
NYANZA	**13.8**	**1.2**	**6.3**	**85**	**895**	**872**
Kisii	21.5	2.6	11.3	25	245	218
Kisumu	17.8	2.8	6.5	9	194	255
Siaya	8.5	2.2	22.9	25	166	134
South Nyanza	6.1	1.4	16.9	26	290	265
RIFT VALLEY	**29.6**	**1.7**	**10.5**	**76**	**742**	**1047**
Baringo	12.5	a	a	2	16	51
Elgeyo Marak	16.7	a	a	2	12	38
Kajiado	52.0	a	a	2	25	80
Kericho	24.1	2.6	12.0	26	277	205
Laikipia	68.5	a	a	3	25	57
Nakuru	47.2	5.4	3.7	9	88	156
Nandi	16.7	a	a	2	30	95
Narok	22.6	a	a	2	31	99
Trans Nzoia	28.9	a	a	2	45	144
Uasin Gishu	13.4	2.5	19.8	24	180	80
West Pokot	0.0	a	a	2	13	42
WESTERN	**13.7**	**1.3**	**15.9**	**54**	**745**	**711**
Bungoma	9.4	1.5	17.8	26	359	185
Busia	16.1	a	a	3	51	119
Kakamega	14.9	2.0	13.4	25	335	407
TOTAL	26.9	1.0	3.6	438	4778	4765

[a] Sampling errors not calculated because fewer than 9 PSUs.

4 Synthetic Estimation

One of the drawbacks of the direct method just described is that the estimate is affected by the large sampling error if the sample size is small. This problem, in turn, inhibits comparison of estimates between areas. One of the ways to overcome the problem of small sample size is to use the proportion of married women from an external source such as a census or a large survey as an adjustment weight. In the literature, several methods of *synthetic estimation* have been discussed (see Aliaga and Le, 1991). The standard synthetic estimation procedure, as suggested in Gonzalez (1973) is used here. The method assumes that the estimated rate is constant within larger areas. What is different between the larger area and the smaller areas within it is the proportion of individuals in the smaller area who have a selected characteristic that is related to the estimate. In other words, in the case of contraceptive prevalence all of the variation in the estimate across the small areas is assumed to be due to differential composition of the populations, not a differential propensity to use contraception. The standard synthetic estimation procedure can be expressed as follows:

$$P_{i.} = \sum_j \left(\frac{N_{ij}}{N_{i.}} \times p_j\right)$$

where:

$P_{i.}$ = the estimated contraceptive prevalence rate in the *i*th administrative area;

p_j = the observed prevalence rate in the *j*th auxiliary category;

N_{ij} = the estimated number of married women age 15-49 in the *j*th auxiliary category for the *i*th area, as observed in the external source;

$N_{i.}$ = the estimated number of married women age 15-49 in the *i*th area, as observed in the external source; and

$N_{ij}/N_{i.}$ = the adjustment weights.

Contraceptive use age patterns were examined for the Dominican Republic and Kenya. The results indicate that the contraceptive prevalence rate increases with an increase in women's age, particularly from age group 15-19 to age group 35-39. Logistic regression analysis also shows that the log odds of contraceptive use are higher for women who are age 25-34 and 35-49 than for those women who are in the 15-24 age group. Since a woman's age is an important predictor of contraceptive use and data on women's age are available from an external source, it was decided to use women's age as a weighting factor in the estimation of the prevalence rate for small population areas. Data on women's age as an auxiliary variable are taken from the 1991 15,000-household survey for the Dominican Republic and from the 1990 Population Census for Kenya.

The distribution of married women in the three age groups (15-24, 25-34, and 35-49) within each province/district was calculated. Homogeneity of prevalence rates is assumed for each age group within the region/province. The estimates produced by the synthetic method are presented in Table 4.1 for the Dominican Republic and in Table 4.2 for Kenya.

Monte Plata (in Region I) in the Dominican Republic is used as an example to show how the calculation is performed. The adjustment weights ($N_{ij}/N_{i.}$) are computed from the auxiliary age data as .29, .32, and .40 for age groups 15-24, 25-34, and 35-49, respectively; the region-level contraceptive prevalence rates are .288, .627, and .590, respectively, for the three age groups. Using the above formula, we estimate the contraceptive prevalence rate as 51.4 percent for the Monte Plata province: (.29 x .288) + (.32 x .627) + (.40 x .590).

In this calculation, the assumption is made that the provinces/districts within the region/province are homogeneous with respect to age-specific contraceptive prevalence rates. Even if the assumption of homogeneity is not correct, any possible extreme estimate is brought in line with the regional estimate. It is observed from Table 4.1, that the province-level estimates are close to regional estimates for the Dominican Republic. A similar situation occurs among districts in Kenya (see Table 4.2), although dissimilarities are stronger among provinces in Kenya than among regions in the Dominican Republic. Because of the fact that the sampling frame for the auxiliary age variable is different from that for the contraceptive use variable, the sampling variance of the synthetic estimate has not been calculated.

Table 4.1 Synthetic estimates of contraceptive prevalence for married women: Dominican Republic 1991

Region and province	Distribution of married women by age[1]			Total percent	Regional contraceptive prevalence rate and relative error[2]				Synthetic estimate of contraceptive prevalence
	15-24	25-34	35-49		15-24 (RE)	25-34 (RE)	35-49 (RE)[3]	15-49	
REGION 0					**36.0 (14.4)**	**60.8 (5.3)**	**73.2 (4.0)**	**60.7**	
D. Nacional	20.0	41.2	38.8	100.0					60.6
REGION I					**28.8 (15.1)**	**62.7 (5.6)**	**59.0 (6.0)**	**50.8**	
Peravia	28.8	32.1	39.2	100.0					51.5
San Cristobal	22.2	37.4	40.5	100.0					53.7
Monte Plata	28.9	31.5	39.7	100.0					51.4
REGION II					**42.3 (11.0)**	**67.0 (4.5)**	**70.0 (5.4)**	**61.0**	
Santiago	24.6	37.0	38.5	100.0					62.1
Puerto Plata	24.2	43.0	32.9	100.0					62.0
La Vega	26.5	38.9	34.6	100.0					61.5
Espaillat	24.7	40.0	35.4	100.0					62.0
Monseñor Nouel	28.0	37.2	34.9	100.0					61.2
REGION III					**39.0 (19.1)**	**65.7 (10.0)**	**60.2 (6.7)**	**57.4**	
Salcedo	25.0	39.9	35.2	100.0					57.1
Duarte	27.9	33.6	38.6	100.0					56.1
M. T. Sanchez	22.3	40.9	36.9	100.0					57.8
Samana	31.4	26.9	41.8	100.0					55.0
S. Ramirez	19.3	43.3	37.4	100.0					58.5
REGION IV					**26.4 (15.6)**	**56.9 (6.1)**	**56.1 (8.5)**	**47.1**	
Barahona	28.5	31.8	39.8	100.0					47.9
Pedernales	38.0	37.2	24.8	100.0					45.1
Bahoruco	29.3	36.9	33.9	100.0					47.7
Independencia	31.5	28.3	40.3	100.0					46.2
REGION V					**33.5 (13.5)**	**51.1 (8.4)**	**66.0 (7.2)**	**50.6**	
La Romana	31.1	39.5	29.5	100.0					50.1
La Altagracia	28.7	35.5	35.9	100.0					51.4
El Seibo	28.3	38.9	32.9	100.0					51.0
S. P. Macoris	24.2	38.8	37.0	100.0					52.3
Hato Mayor	23.2	43.5	33.3	100.0					51.9
REGION VI					**22.8 (27.3)**	**46.8 (10.0)**	**45.1 (17.4)**	**39.7**	
San Juan	27.8	36.6	35.6	100.0					39.5
Azua	29.6	33.9	36.5	100.0					39.0
Elías Piña	26.0	33.0	41.1	100.0					39.9
REGION VII					**26.2 (16.6)**	**74.3 (7.3)**	**65.3 (7.9)**	**58.6**	
Valverde	24.2	36.6	39.3	100.0					59.1
S. Rodriguez	15.5	38.6	45.9	100.0					62.7
Dajabon	29.9	32.9	37.2	100.0					56.5
Monte Cristi	30.3	30.6	39.2	100.0					56.3

Note: Values in parentheses indicate relative errors.
[1]1991 Expanded Household Survey (IEPD, ONAPLAN, and MI, 1993)
[2]1991 DHS survey (IEPD, ONAPLAN, and IRD, 1992)
[3]See Table 3.1 for relative error

Table 4.2 Synthetic estimates of contraceptive prevalence for married women: Kenya, 1989

Province and district	Distribution of married women by age[1]			Total	Regional contraceptive prevalence rate and relative error[2]				Synthetic estimate of contraceptive prevalence
	15-24	25-34	35-49	percent	15-24 (RE)	25-34 (RE)	35-49 (RE)[3]	15-49	
NAIROBI					**19.1 (15.9)**	**39.2 (10.0)**	**49.1(8.5)**	**33.5**	
Nairobi	32.7	44.8	22.5	100.0					34.9
CENTRAL					**28.2 (21.6)**	**39.1(7.0)**	**45.2(8.9)**	**39.5**	
Kiambu	27.1	40.1	32.8	100.0					38.1
Kirinyaga	25.1	39.5	35.4	100.0					38.5
Murang'a	21.6	38.8	39.5	100.0					39.2
Nyandarua	26.8	37.6	35.6	100.0					38.4
Nyeri	22.3	38.9	38.8	100.0					39.0
COAST					**6.9 (39.7)**	**23.9 (12.5)**	**18.6 (20.8)**	**18.1**	
Kilifi	29.3	37.8	32.9	100.0					17.2
Kwale	30.7	38.2	31.1	100.0					17.0
Mombasa	32.5	42.2	25.4	100.0					17.0
Taita	22.4	39.1	38.4	100.0					18.1
EASTERN					**32.4 (13.0)**	**43.6(8.8)**	**40.5(9.8)**	**40.2**	
Embu	21.9	41.0	37.2	100.0					40.0
Kitui	25.1	36.4	38.5	100.0					39.6
Machakos	21.6	40.3	38.1	100.0					40.0
Meru	22.8	41.4	35.8	100.0					39.9
NYANZA					**9.9 (24.5)**	**13.9(9.0)**	**16.2 (15.4)**	**13.8**	
Kisii	27.9	42.2	30.0	100.0					13.5
Kisumu	31.4	38.7	29.9	100.0					13.3
Siaya	26.5	35.5	38.0	100.0					13.7
South Nyanza	34.6	36.4	29.0	100.0					13.2
RIFT VALLEY					**25.6 (20.9)**	**30.2 (12.9)**	**31.4 (17.2)**	**29.6**	
Baringo	26.5	40.6	33.0	100.0					29.4
Elgeyo Marak	26.5	39.2	34.3	100.0					29.4
Kajiado	38.1	36.3	25.6	100.0					28.8
Kericho	31.7	39.3	29.1	100.0					29.1
Laikipia	26.6	39.1	34.3	100.0					29.4
Nakuru	30.5	39.9	29.6	100.0					29.2
Nandi	28.1	39.1	32.8	100.0					29.3
Narok	37.4	36.5	26.1	100.0					28.8
Trans Nzoia	29.8	39.8	30.5	100.0					29.2
Uasin Gishu	29.7	40.1	30.2	100.0					29.2
West Pokot	32.7	37.5	29.8	100.0					29.1
WESTERN					**5.7 (32.0)**	**15.6 (20.8)**	**17.6 (18.9)**	**13.7**	
Bungoma	30.7	39.6	29.7	100.0					13.2
Busia	30.9	37.4	31.7	100.0					13.2
Kakamega	U	U	U	U					U

Note: Values in parentheses indicate relative errors.
U=Unknown (not available)
[1]1990 Population Census
[2]1989 DHS survey (NCPD and IRD, 1989)
[3]See Table 3.2 for relative error

5 Regression Method

One of the main weaknesses of the synthetic method is that only a few auxiliary variables can be used at a time. Another problem is that the sampling and nonsampling errors of the auxiliary variables from the external source are not compatible with those of the prevalence estimates from the main source. Using the *regression approach* alleviates these problems. The regression method includes all the potential predictors related to current use of contraception, while relying only on data from the main source.

The DHS data from both the Dominican Republic and Kenya suggest that use of contraception is associated with a woman's age, education, number of living children, work status, floor type, urban-rural residence, water supply, electricity in the house, possession of a radio, toilet type, and husband's education. These 11 variables are reformulated into 11 dummy variables (1,0), based on the cutoffs determined through a two-step process to maximize the power of discrimination of the predictor (Goldstein and Dillon, 1978). First, the cumulative distribution of each of the predictors is constructed for two subgroups: those who are currently using a family planning method and those who are not. Then, the absolute difference in the cumulative frequency between the two subgroups is calculated for each of the categories of the predictor. The cutoff point, x^*, is that value (or the category) of the original predictor which maximizes the difference in the cumulative frequency between use and nonuse subgroups, which can be expressed as follows:

$$|F_u(x^*) - F_n(x^*)| = \max_x |F_u(x) - F_n(x)|$$

where:

$F_u(x)$ = the cumulative frequency distribution of the original predictor at value x for the use subgroup; and

$F_n(x)$ = the cumulative frequency distribution of the original predictor at value x for the nonuse subgroup.

The 11 predictor dummies and the dummy variable on current contraception (1,0) are constructed at the individual level. The values of these 11 dummies are averaged in order to create a cluster-level numeric variable, expressed in terms of a proportion ($p_1 ... p_{11}$). Similarly, the values of the contraceptive use dummy are converted into the prevalence rate, P_c, at the cluster level. This is done from cluster 1 to m (where m is the total number of clusters), as shown below. As noted earlier, the cluster is a sample of selected households in each PSU.

Cluster No.	p_1	p_2	...	p_{11}	P_c
1					
2					
3					
4					
.					
.					
.					
m					

The contraceptive prevalence rate is regressed on the 11 factors, with the cluster as the unit of analysis. The regression equation can be expressed as follows:

$$P_c = \alpha_0 + \alpha_1 p_{1c} + \alpha_2 p_{2c} + ... + \alpha_{11} p_{11c}$$

where:

P_c = the observed prevalence rate in cluster c (=1,..., m);
α_0 = the intercept;
α_1 = the coefficient for predictor p_1;
α_2 = the coefficient for predictor p_2;
α_{11} = the coefficient for predictor p_{11}; and
e_c = the error term for cluster c.

Using the backward selection option of the linear regression procedure in the Statistical Package for Social Sciences (SPSS), the best model is identified. No sampling weights were used in the regression analysis; however, the predicted prevalence rate at the cluster-level estimated from the regression equation is adjusted for sampling weights to produce the final province/district level estimates, using the following formula:

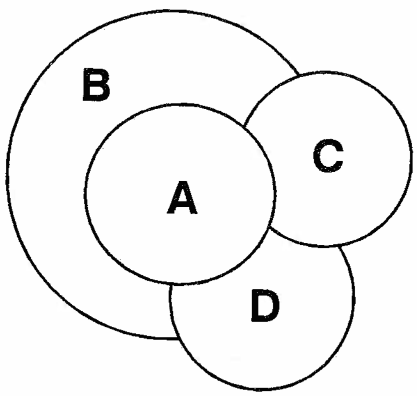
$$P = \frac{\Sigma b_c w_c P_{pc}}{\Sigma b_c w_c}$$

where:

P = the prevalence rate at the province/district level;
b_c = the number of women in cluster c;
w_c = the sampling weight for cluster c;
P_{pc} = the predicted prevalence rate for cluster c; and

the sum Σ is over all clusters in the given province/district.

If a given province in the Dominican Republic or a district in Kenya has fewer than 30 clusters, several new sets of clusters were formed by adding clusters from the contiguous province/district such that some original clusters remain in each of the new sets. The predicted prevalence rates estimated from the models for the new sets are averaged. This average rate is the final province-level prevalence estimate. This can be illustrated by the following Venn Diagram:



The diagram has four subsets of clusters representing Provinces A, B, C, and D. The objective is to estimate the contraceptive prevalence rate for Province A, which has fewer than 30 clusters. It is necessary to add clusters from the neighboring provinces to meet the required number of at least 30 clusters. There exist at least three possible groupings: A and B; A, C, and D, and A, D, and B and the corresponding prevalence rates are *P(I)*, *P(II)* and *P(III)*. The estimated rate for Province A is obtained by averaging the three rates.

Three sets of estimates are presented for the Dominican Republic in Table 5.1 and for Kenya in Table 5.2. Note that there are some regions in the Dominican Republic (e.g., Region VI) and some provinces in Kenya (e.g., Rift Valley) which have provincial/district prevalence estimates that vary dramatically. These extreme values have been smoothed out (are brought in line with regional variation) by the regression method. However, the interprovince variability (for the Dominican Republic) or the interdistrict variability (for Kenya) is only slightly reduced compared to the direct estimate. In contrast, the interarea variability in the estimate obtained from the synthetic method is reduced considerably. Tables 5.1 and 5.2 also show that the estimates of the sampling errors in the regression method are, in general, lower than those in the direct method.

Table 5.1 Direct, synthetic, and regression estimates of contraceptive prevalence for married women: Dominican Republic, 1991

Region and province	Direct estimate		Synthetic estimate	Regression estimate	
	Value	Sampling error		Value	Sampling error
REGION 0	**60.7**				
Distrito Nacional	60.7	2.0	60.6	61.7	1.8
REGION I	**50.8**				
Peravia	48.7	5.8	51.5	54.6	3.0
San Cristóbal	55.9	3.7	53.7	52.3	2.6
Monte Plata	43.1	3.9	51.4	52.3	3.5
REGION II	**61.0**				
Santiago	62.1	3.5	62.1	60.7	4.4
Puerto Plata	62.8	5.3	62.0	58.6	2.1
La Vega	59.4	4.4	61.5	58.0	2.0
Espaillat	56.2	a	62.0	57.3	3.5
Monseñor Nouel	61.7	a	61.2	56.7	2.0
REGION III	**57.4**				
Salcedo	53.4	a	57.1	56.1	3.4
Duarte	57.2	4.5	56.1	55.3	2.3
María T. Sánchez	65.9	7.5	57.8	57.9	1.6
Samaná	44.6	a	55.0	57.9	1.6
Sanchez Ramirez	56.7	8.2	58.5	53.7	4.3
REGION IV	**47.1**				
Barahona	51.9	2.7	47.9	50.0	3.3
Pedernales	48.3	a	45.1	50.0	2.4
Bahoruco	38.8	6.5	47.7	43.4	4.8
Independencia	44.8	a	46.2	49.1	2.5
REGION V	**50.6**				
La Romana	44.0	7.9	50.1	50.1	4.0
La Altagracia	57.5	a	51.4	49.8	4.1
El Seybo	46.5	a	51.0	49.5	3.4
San P. de Macorís	54.2	3.5	52.3	50.8	3.3
Hato Mayor	58.0	a	51.9	49.6	2.7
REGION VI	**39.7**				
San Juan	34.1	8.2	39.5	41.5	6.6
Azua	52.3	4.5	39.0	52.7	3.2
Elías Piña	32.1	a	39.9	40.5	8.0
REGION VII	**58.6**				
Valverde	53.9	2.4	59.1	58.4	2.9
Santiago Rodriguez	73.6	a	62.7	55.1	4.9
Dajabón	51.8	a	56.5	55.8	6.3
Monte Cristi	62.8	4.4	56.3	59.3	2.6

[a] Sampling errors not calculated because fewer than 9 PSUs.

Table 5.2 Direct, synthetic, and regression estimates of contraceptive prevalence for married women: Kenya, 1989

Province and district	Direct estimate		Synthetic estimate	Regression estimate	
	Value	Sampling error		Value	Sampling error
NAIROBI					
Nairobi	33.5	2.1	34.9	38.5	3.6
CENTRAL	**39.5**				
Kiambu	37.3	a	38.1	37.9	3.2
Kirinyaga	54.2	3.2	38.5	44.4	2.0
Murang'a	32.1	5.4	39.2	38.8	2.5
Nyandarua	39.2	a	38.4	40.4	3.2
Nyeri	40.8	3.2	39.0	43.3	2.8
COAST	**18.1**				
Kilifi	10.8	1.7	17.2	15.2	2.6
Kwale	16.2	a	17.0	17.3	4.3
Mombasa	24.5	3.6	17.0	15.0	2.0
Taita	30.7	a	18.1	29.0	1.9
EASTERN	**40.2**				
Embu	47.2	a	40.0	42.1	2.4
Kitui	41.3	a	39.6	41.6	2.7
Machakos	40.4	2.9	40.0	41.0	1.7
Meru	36.3	3.5	39.9	43.4	3.8
NYANZA	**13.8**				
Kisii	21.5	2.6	13.5	16.9	1.4
Kisumu	17.8	2.8	13.3	14.8	1.3
Siaya	8.5	2.2	13.7	10.1	1.5
South Nyanza	6.1	1.4	13.2	11.2	1.0
RIFT VALLEY	**29.6**				
Baringo	12.5	a	29.4	27.1	3.1
Elgeyo Marak	16.7	a	29.4	15.0	2.8
Kajiado	52.0	a	28.8	41.5	4.6
Kericho	24.1	2.6	29.1	19.3	1.6
Laikipia	68.5	a	29.4	45.0	4.5
Nakuru	47.2	5.4	29.2	34.0	5.2
Nandi	16.7	a	29.3	18.0	2.2
Narok	22.6	a	28.8	23.3	4.5
Trans Nzoia	28.9	a	29.2	15.1	2.5
Uasin Gishu	13.4	2.5	29.2	15.7	2.3
West Pokot	0.0	a	29.1	U	U
WESTERN	**13.7**				
Bungoma	9.4	1.5	13.2	11.8	1.8
Busia	16.1	a	13.2	11.7	2.2
Kakamega	14.9	2.0	U	14.4	1.7

U=Unknown (data required for estimate were not available)
[a] Sampling errors not calculated because fewer than 9 PSUs.

6 Conclusions

The estimates provided by these three methods are consistent (the larger the number of PSUs in the area, the better the estimate). The range of both synthetic and regression estimates for provinces/districts within each region in the Dominican Republic and each province in Kenya is smaller than the range for the direct estimate (see Tables 6.1 and 6.2). For the Dominican Republic, the correlation value between the direct and synthetic provincial estimates is 0.90, 0.88 between the direct and regression provincial estimates, and 0.95 between the synthetic and regression provincial estimates. For the 30 districts in Kenya (excluding West Pokot and Kakamega), the correlation values between district estimates are 0.70, 0.90, and 0.85, respectively. Restricting the analysis to provinces/districts having at least 200 cases yields slightly higher correlation values, supporting the finding that the estimates are consistent. In general, the correlation values are slightly higher for the Dominican Republic than for Kenya, probably due to the fact that provinces in the Dominican Republic are more homogenous than districts in Kenya.

Of the three methods used to estimate contraceptive prevalence rates, the regression approach was found most suitable. First, the approach overcomes the problem of small numbers of cases encountered in the direct estimation procedure by averaging the estimates of several sets, each of which contains the common areas (provinces of Dominican Republic and districts of Kenya) of interest. Although this could be done with the direct method, there are no objective criteria for selecting which areas should form a set. Such criteria in the regression approach are developed by finding whether the same set of predictors explains the use of contraception in the areas that constitute a set. Second, unlike the synthetic approach, no auxiliary data are needed in the regression approach and the assumption of homogeneity is not required. Third, the regression method allows for estimates of sampling error, while the synthetic procedure does not.

Another possible approach to indirect estimation that is not examined in this report is the empirical Bayes approach (Fay and Herriot, 1979), which would essentially be a weighted average of the direct province/district estimate and the regional/provincial estimate of contraceptive prevalence. This new prevalence estimate gives more weight to the region/province estimate (more accurate) than to the province/district estimate.

Table 6.1 Range of contraceptive prevalence rates among provinces by region, midpoint ± half range: Dominican Republic, 1991

Region	Direct estimate	Synthetic estimate	Regression estimate
Region I	49.50 ± 6.40	52.55 ± 1.15	53.45 ± 1.15
Region II	59.50 ± 3.30	61.65 ± 0.45	58.70 ± 2.00
Region III	55.25 ± 10.65	56.75 ± 1.25	55.80 ± 2.10
Region IV	45.35 ± 6.55	46.50 ± 1.40	46.70 ± 3.30
Region V	51.00 ± 7.00	51.20 ± 1.10	50.15 ± 0.65
Region VI	42.20 ± 10.10	39.45 ± 0.45	46.60 ± 6.10
Region VII	62.70 ± 10.90	59.50 ± 3.20	57.20 ± 2.10
TOTAL	52.85 ± 20.75	50.85 ± 11.85	51.10 ± 10.60

Table 6.2 Range of contraceptive prevalence rates among districts by province, midpoint ± half range: Kenya, 1989

Province	Direct estimate	Synthetic estimate	Regression estimate
Central	43.15 ± 11.05	38.65 ± 0.55	41.15 ± 3.25
Coast	20.75 ± 9.95	17.55 ± 0.55	22.00 ± 7.00
Eastern	41.75 ± 5.45	39.80 ± 0.20	42.20 ± 1.20
Nyanza	13.80 ± 7.70	13.45 ± 0.25	13.50 ± 3.40
Rift Valley	40.50 ± 28.00	29.10 ± 0.30	30.00 ± 15.00
Western	12.75 ± 3.35	13.20 ± 0.00	13.05 ± 1.35
TOTAL	37.30 ± 31.20	26.60 ± 13.40	27.55 ± 17.45

References

Aliaga, A. and T. Le. 1991. Methodology for Small Area Estimation with DHS Surveys. *Proceedings of the Demographic and Health Surveys World Conference Washington, D.C.*, 1991. Vol. 1. pp. 497-512. Columbia, Maryland: IRD/Macro International.

Fay, R.E. and R.A. Herriot. 1979. Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association* 74:269-277.

Goldstein, M. and N. Dillon. 1978. *Discrete Discriminant Analysis*. pp. 70-71. New York: John Wiley.

Gonzalez, M.E. 1973. Use and Evaluation of Synthetic Estimates. *Proceedings of the American Statistical Association* (Social Statistics Section), December 27-30. pp. 33-36. Washington, D.C.: ASA.

Instituto de Estudios de Población y Desarrollo (IEPD) and Oficina Nacional de Planificación (ONAPLAN) and IRD/Macro International Inc. (IRD). 1992. *Encuesta Demográfica y de Salud 1991*. Columbia, Maryland: IEDP and IRD.

Instituto de Estudios de Población y Desarrollo (IEPD) and Oficina Nacional de Planificación (ONAPLAN) and Macro International Inc. (MI). 1993. *Encuesta Demográfica y de Salud 1991. Resultados Generales Cuestionario Hogar Ampliado*. Santo Domingo, Dominican Republic.

Institute for Resource Development. 1987. *Sampling Manual*. DHS Basic Documentation No. 8. Columbia, Maryland.

Muhuri, P.K. and S.O. Rutstein. 1994. *Socioeconomic, Demographic, and Health Indicators for Subnational Areas*. DHS Comparative Studies No. 9. Calverton, Maryland: Macro International Inc.

National Council for Population and Development (NCPD) and Institute for Resource Development/Macro Systems Inc. (IRD). 1989. *Kenya Demographic and Health Survey 1989*. Columbia, Maryland: NCPD and IRD.

Appendix

Summary of DHS-I and DHS-II Surveys, 1985-1993

Region and Country	Date of Fieldwork		Implementing Organization	Respondents	Sample Size	Male/Husband Survey	Supplemental Studies, Modules, and Additional Questions
SUB-SAHARAN AFRICA							
DHS-I							
Botswana	Aug-Dec	1988	Central Statistics Office	AW 15-49	4,368		AIDS, PC, adolescent fertility
Burundi	Apr-Jul	1987	Département de la Population, Ministère de l'Intérieur	AW 15-49	3,970	542 Husbands	CA, SAI, adult mortality
Ghana	Feb-May	1988	Ghana Statistical Service	AW 15-49	4,488	943 Husbands	CA, SM, WE
Kenya	Dec-May	1988/89	National Council for Population and Development	AW 15-49	7,150	1,133 Husbands	
Liberia	Feb-Jul	1986	Bureau of Statistics, Ministry of Planning and Economic Affairs	AW 15-49	5,239		TBH, employment status
Mali	Mar-Aug	1987	Institut du Sahel, USED/CERPOD	AW 15-49	3,200	970 Men 20-55	CA, VC, childhood physical handicaps
Ondo State, Nigeria	Sep-Jan	1986/87	Ministry of Health, Ondo State	AW 15-49	4,213		CA, TBH
Senegal	Apr-Jul	1986	Direction de la Statistique, Ministère de l'Economie et des Finances	AW 15-49	4,415		CA, CD
Sudan	Nov-May	1989/90	Department of Statistics, Ministry of Economic and National Planning	EMW 15-49	5,860		M, MM, female circumcision, family planning services
Togo	Jun-Nov	1988	Unité de Recherche Démographique, Université du Benin	AW 15-49	3,360		CA, SAI, marriage history
Uganda	Sep-Feb	1988/89	Ministry of Health	AW 15-49	4,730		CA, SAI
Zimbabwe	Sep-Jan	1988/89	Central Statistical Office	AW 15-49	4,201		AIDS, CA, PC, SAI, WE
DHS-II							
Burkina Faso	Dec-Mar	1992/93	Institut National de la Statistique et de la Démographie	AW 15-49	6,000	1,845 Men 18+	AIDS, CA, MA, SAI
Cameroon	Apr-Sep	1991	Direction Nationale du Deuxiême Recensement Général de la Population et de l'Habitat	AW 15-49	3,871	814 Husbands	CA, CD, SAI
Madagascar	May-Nov	1992	Centre National de Recherches sur l'Environnement	AW 15-49	6,260		CA, MM, SAI
Malawi	Sep-Nov	1992	National Statistical Office	AW 15-49	4,850	1,151 Men 20-54	AIDS, CA, MA, MM, SAI
Namibia	Jul-Nov	1992	Ministry of Health and Social Services, Central Statistical Office	AW 15-49	5,421		CA, CD, MA, MM
Niger	Mar-Jun	1992	Direction de la Statistique et des Comtes Nationaux	AW 15-49	6,503	1,570 Husbands	CA, MA, MM, SAI
Nigeria	Apr-Oct	1990	Federal Office of Statistics	AW 15-49	8,781		CA, SAI
Rwanda	Jun-Oct	1992	Office National de la Population	AW 15-49	6,551	598 Husbands	CA, SAI
Senegal	Nov-Aug	1992/93	Direction de la Prévision et de la Statistique	AW 15-49	6,310	1,436 Men 20+	AIDS, CA, MA, MM, SAI
Tanzania	Oct-Mar	1991/92	Bureau of Statistics, Planning Commission	AW 15-49	9,238	2,114 Men 15-60	AIDS, CA, MA, SAI
Zambia	Jan-May	1992	University of Zambia	AW 15-49	7,060		AIDS, CA, MA
NEAR EAST/NORTH AFRICA							
DHS-I							
Egypt	Oct-Jan	1988/89	National Population Council	EMW 15-49	8,911		CA, CD, MM, PC, SAI, WE, women's status
Morocco	May-Jul	1987	Ministère de la Santé Publique	EMW 15-49	5,982		CA, CD, S
Tunisia	Jun-Oct	1988	Office National de la Famille et de la Population	EMW 15-49	4,184		CA, CD, S, SAI
DHS-II							
Egypt	Nov-Dec	1992	National Population Council	EMW 15-49	9,864	2,406 Husbands	CA, MA, PC, SM
Jordan	Oct-Dec	1990	Department of Statistics, Ministry of Health	EMW 15-49	6,462		CA, SAI
Morocco	Jan-Apr	1992	Ministère de la Santé Publique	AW 15-49	9,256	1,336 Men 20-70	CA, MA, MM, SAI
Yemen	Nov-Jan	1991/92	Central Statistical Organization	EMW 15-49	5,687		CA, CD, SAI

Region and Country	Date of Fieldwork		Implementing Organization	Respondents	Sample Size	Male/Husband Survey	Supplemental Studies, Modules, and Additional Questions
ASIA							
DHS-I							
Indonesia	Sep-Dec	1987	Central Bureau of Statistics, National Family Planning Coordinating Board	EMW 15-49	11,844		PC, SM
Nepal (In-depth)	Feb-Apr	1987	New Era	CMW 15-49	1,623		KAP–gap survey
Sri Lanka	Jan-Mar	1987	Department of Census and Statistics, Ministry of Plan Implementation	EMW 15-49	5,865		CA, NFP
Thailand	Mar-Jun	1987	Institute of Population Studies, Chulalongkorn University	EMW 15-49	6,775		CA, S, SAI
DHS-II							
Indonesia	May-Jul	1991	Central Bureau of Statistics, National Family Planning Coordinating Board, Ministry of Health	EMW 15-49	22,909		PC, SM
Pakistan	Dec-May	1990/91	National Institute of Population Studies	EMW 15-49	6,611	1,354 Husbands	CA
LATIN AMERICA & CARIBBEAN							
DHS-I							
Bolivia	Mar-Jun	1989	Instituto Nacional de Estadística	AW 15-49	7,923		CA, CD, MM, PC, S, WE
Bolivia (In-depth)	Mar-Jun	1989	Instituto Nacional de Estadística	AW 15-49	7,923		Health
Brazil	May-Aug	1986	Sociedade Civil Bem-Estar Familiar no Brasil	AW 15-44	5,892		CA, PC, SM, abortion, young adult use of contraception
Colombia	Oct-Dec	1986	Corporación Centro Regional de Población, Ministerio de Salud	AW 15-49	5,329		CA, PC, SAI, SM
Dominican Republic	Sep-Dec	1986	Consejo Nacional de Población y Familia	AW 15-49	7,649		NFP, S, SAI, SM family planning communication
Dominican Rep. (Experimental)	Sep-Dec	1986	Consejo Nacional de Población y Familia	AW 15-49	3,885		
Ecuador	Jan-Mar	1987	Centro de Estudios de Población y Paternidad Responsable	AW 15-49	4,713		CD, SAI, employment
El Salvador	May-Jun	1985	Asociación Demográfica Salvadoreña	AW 15-49	5,207		S, TBH
Guatemala	Oct-Dec	1987	Instituto de Nutrición de Centro América y Panamá	AW 15-44	5,160		S, SAI
Mexico	Feb-May	1987	Dirección General de Planificación Familiar Secretaría de Salud	AW 15-49	9,310		NFP, S, employment
Peru	Sep-Dec	1986	Instituto Nacional de Estadística	AW 15-49	4,999		NFP, employment, cost of family planning
Peru (Experimental)	Sep-Dec	1986	Instituto Nacional de Estadística	AW 15-49	2,534		
Trinidad and Tobago	May-Aug	1987	Family Planning Association of Trinidad and Tobago	AW 15-49	3,806		CA, NFP, breastfeeding
DHS-II							
Brazil (NE)	Sep-Dec	1991	Sociedade Civil Bem-Estar Familiar no Brasil	AW 15-49	6,222	1,266 Husbands	AIDS, PC
Colombia	May-Aug	1990	PROFAMILIA	AW 15-49	8,644		AIDS
Dominican Republic	Jul-Nov	1991	Instituto de Estudios de Población y Desarrollo (PROFAMILIA), Oficina Nacional de Planificación	AW 15-49	7,320		CA, MA, S, SAI
Paraguay	May-Aug	1990	Centro Paraguayo de Estudios de Población	AW 15-49	5,827		CA, SAI
Peru	Oct-Mar	1991/92	Instituto Nacional de Estadística e Informática	AW 15-49	15,882		CA, MA, MM, SAI

AW all women
CMW currently married women
EMW ever-married women

AIDS acquired immune deficiency syndrome
CA child anthropometry
CD causes of death (verbal reports of symptoms)
M migration
MA maternal anthropometry
MM maternal mortality

NFP natural family planning
PC pill compliance
S sterilization
SAI service availability information
SM social marketing
TBH truncated birth history
VC value of children