

LOS DISEÑOS MUESTRALES UTILIZADOS EN LA AMERICA  
LATINA EN LA ENCUESTA MUNDIAL DE FECUNDIDAD:  
ALGUNOS DETERMINANTES Y CONSECUENCIAS \*

*Alberto M. Marckwardt*  
(CELADE)

RESUMEN

Se describen, en forma resumida, las características de los diseños muestrales utilizados en los nueve países latinoamericanos que han participado en el programa de la Encuesta Mundial de la Fecundidad. Luego se presentan algunos errores muestrales representativos de cuatro países, y las conclusiones que de éstos se derivan. Finalmente se presenta un resumen de los resultados principales de un estudio de errores no muestrales, llevado a cabo en el Perú.

<ENCUESTA MUNDIAL DE FECUNDIDAD> <MUESTREO> <ERROR DE MUESTREO>

\* Este trabajo fue presentado al Seminario Regional sobre Utilización de Muestreo en Encuestas de Hogares, Santiago de Chile, 27 de octubre al 7 de noviembre de 1980.

SAMPLING DESIGNS USED IN LATIN AMERICAN  
COUNTRIES IN THE WORLD FERTILITY SURVEY:  
DETERMINANTS AND CONSEQUENCES

SUMMARY

The paper describes in summary form the characteristics of sampling designs used in the nine Spanish-speaking Latin American countries that have participated in the programme of the World Fertility Survey. It goes on to present some representative sampling errors from four countries, and the conclusions derived therefrom. Finally, the principal results of a study of non-sampling errors carried out in Peru are summarized.

<WORLD FERTILITY SURVEY>    <SAMPLING>  
<SAMPLING ERROR>

## 1. INTRODUCCION

La Encuesta Mundial de Fecundidad (EMF) es un programa internacional de investigación cuyo propósito es determinar el estado actual de la fecundidad humana en el mundo. Para lograr este objetivo, se están promoviendo y financiando encuestas de fecundidad por muestreo en el mayor número posible de países. Estas encuestas son diseñadas y realizadas científicamente, y además son nacionalmente representativas y comparables a nivel internacional.

El proyecto está a cargo del Instituto Internacional de Estadística, contando con la colaboración de la División de Población de las Naciones Unidas y en cooperación con la Unión Internacional para el Estudio Científico de la Población. Es financiado principalmente por el Fondo de las Naciones Unidas para Actividades de Población y por la Agencia para el Desarrollo Internacional de los Estados Unidos. La Oficina Británica para el Desarrollo de Países Extranjeros proporciona también un gran apoyo financiero.

La EMF estableció su sede en Londres, e inició sus actividades en 1972; el trabajo de campo empezó en 1974. A mediados de 1980, un total de 43 países en vías de desarrollo y 22 países desarrollados habían participado en el programa. (Los países desarrollados financian sus propias encuestas). Entre los 43 países en vías de desarrollo, 37 ya han completado su trabajo de campo. La distribución geográfica de estos países es la siguiente: 8 en África, 11 en Asia y el Pacífico, 5 en el Oriente Medio, 4 en el Caribe, y 9 en América Latina. Típicamente, la preparación del informe general de resultados de un país requiere un tiempo que varía entre uno y tres años después de completado el trabajo de campo. Hasta la fecha, son 22 los países que han publicado sus informes generales de resultados: 1 en África, 10 en Asia y el Pacífico, 2 en el Oriente Medio, 3 en el Caribe, y 6 en América Latina.

Los países latinoamericanos que han participado en el programa de la EMF son nueve: Colombia, Costa Rica, República Dominicana, Ecuador, México, Panamá, Paraguay, Perú y Venezuela. (Brasil entrará en el programa a fines de 1981). En los nueve, ya se ha completado el trabajo de campo, y los informes de resultados se han publicado en los casos de Colombia, Costa Rica, República Dominicana, México, Panamá y Perú. Los informes de Paraguay y Venezuela se encuentran en imprenta, y el de Ecuador se espera para el primer semestre de 1981.

Los propósitos de este documento son tres:

- a) Describir, en forma resumida, las características de los diseños muestrales utilizados en Latinoamérica en relación al programa de la EMF.
- b) Presentar algunos errores muestrales y las conclusiones que se derivan de éstos; y
- c) Resumir los resultados principales de un estudio de errores no muestrales llevado a cabo en el Perú y en algunos países de otras regiones. Casi todos los resultados citados en este documento han sido ya publicados o presentados en otros foros. Los documentos 1] resumidos en éste son:
  - V. Verma, C. Scott y C. O'Muircheartaigh, *Sample Designs and Sampling Errors for the World Fertility Survey*, ponencia presentada ante la Royal Statistical Society, Londres, 4/1980.
  - V. Verma, *Sampling for National Fertility Surveys*, ponencia preparada para la World Fertility Survey Conference, Londres, 7/80.
  - C. O'Muircheartaigh, A. Marckwardt y V. Verma, *Response Reliability in WFS Data*, ponencia invitada y presentada para la reunión anual No. 42 del Instituto Internacional de Estadística, Manila, 12/79.
  - C. O'Muircheartaigh y A. Marckwardt, *An Assessment of the Reliability of WFS Data*, ponencia preparada para la World Fertility Survey Conference, Londres, 7/80.
  - C. Scott y S. Singh, *Problems of Data Collection in the World Fertility Survey*, ponencia preparada para la World Fertility Survey Conference, Londres, 7/80.

## 2. DISEÑOS MUESTRALES

Desde el comienzo del Programa de la Encuesta Mundial de Fecundidad se ha insistido en el uso de muestras nacionales probabilísticas. El Manual de Diseño Muestral de la EMF declara textualmente: "El muestreo probabilístico es el único enfoque capaz de proporcionar una base científica para generalizar en términos de la población estudiada y un procedimiento para poder estimar el error de muestreo; en el resto del presente manual se da por entendida la necesidad de un muestreo probabilístico estricto." (pág. 2). Otro objetivo ha sido obtener una cobertura máxima, a nivel nacional, en los países participantes.

---

1] Copia de estos trabajos pueden obtenerse escribiendo a la Oficina de Publicaciones, Instituto Internacional de Estadística, Prinses Beatrixlaan 428, Apdo. Postal 950, 2270 AZ Voorburg, Países Bajos.

Como se aprecia en el cuadro 1, los países latinoamericanos han cumplido ampliamente con este objetivo. La cobertura de viviendas particulares ha variado entre el 94 por ciento y el 100 por ciento. Las áreas excluidas son, generalmente, de muy difícil acceso o de escasa densidad poblacional, como en los casos del área rural de la selva en Ecuador, o El Chaco en Paraguay.

Los tamaños de muestra han variado básicamente entre 3.000 y 7.000 mujeres en edad fértil. La decisión sobre el tamaño de muestra depende de tres factores: *a)* el financiamiento disponible; *b)* el nivel de desagregación deseado, particularmente el número de dominios de estudio; y *c)* los intereses analíticos de los directores del estudio. Por ejemplo, en la República Dominicana se decidió utilizar solamente tres áreas de estudio: el área metropolitana de Santo Domingo, el resto urbano, y el área rural. Para esto, bastaba una muestra de 3.000 mujeres. En Ecuador, en cambio, se requerían estimaciones confiables a los siguientes niveles: las regiones económicas 2, 3, 4 y 5; las regiones naturales costa y sierra; niveles urbano y rural; y a nivel de las ciudades de Quito y Guayaquil. Esto determinó una muestra de aproximadamente 7.000 mujeres. En el caso del Perú, los dominios muestrales eran las "Regiones de planificación". Debido a la poca densidad poblacional en la Región Oriente (selva), y el requerimiento de proporcionar por lo menos 1.000 entrevistas en cada región, se habría necesitado una muestra de, por lo menos, 20.000 mujeres en edad fértil. No habiendo financiamiento disponible para una encuesta de ese tamaño, se decidió entrevistar sólo a mujeres alguna vez casadas o unidas (o sea, excluir a las solteras), y aplicar una tasa de muestreo, en la Región Oriente, cuatro veces mayor que en el resto del país. Así se logró bajar el tamaño de muestra requerido a una cifra de entre cinco y seis mil mujeres.

En cuatro de los nueve países latinoamericanos (Colombia, México, Perú y Venezuela), se ha podido hacer uso de un marco muestral maestro ya existente para la encuesta de fecundidad. Este es un hecho alentador, porque refleja el avance general en el campo de las encuestas en la región. Por supuesto, la existencia de dicho marco no soluciona todos los problemas. En primer lugar, es posible que el marco no sea lo más apropiado para una encuesta de fecundidad. En segundo lugar, los listados de vivienda suelen ser muy antiguos. De hecho, en los cuatro países mencionados fue necesario invertir bastante tiempo y dinero en la reactualización de tales listados. Sin embargo, los gastos para una reactualización son menores que el costo de construir una muestra desde cero.

Cuadro 1

*(continúa)*

## CARACTERISTICAS DE LAS MUESTRAS

País	Cobertura (%)	Se basa en marco mues- tral existente	Número de entrevistas	No. de etapas 1 ]
Colombia	99	sí	5.400	U 1 R 2
Costa Rica	97	no	3.900	U 1 R 2
República Dominicana	100	no	3.100	U 2 R 3
Ecuador	96	no	7.000	1
México	100	sí	7.300	U 2 R 3
Panamá	96	no	3.700	U 1 R 2
Paraguay	94	no	4.600	U 1 R 2
Perú	100	sí	5.600	U 1 R 2
Venezuela	98	sí	4.400	1

1 ] Según sea área urbana o rural, excluyendo la etapa de selección de la vivienda.

## CARACTERISTICAS DE LAS MUESTRAS

País	No. UPM 2]	No. UUM 3]	Submuestreo dentro de UUM	Entrevistas por UUM
Colombia	370 35	370 630	1:2,5	5,4
Costa Rica	262 26	262 78	1:6,0	11,5
República Dominicana	26	553 112	1:10,0	4,7
Ecuador	238	238	1:2,0	29,4
México	182	2.280	1:14,4	3,2
Panamá	354 54	354 108	1:3,0	8,0
Paraguay	100 40	100 240	1:5,0	11,8
Perú	410	1.424	1:8,0	3,9
Venezuela	300	300	1:2,0	13,3

2] UPM: unidad primaria de muestro.

3] UUM: unidad última de muestreo.

En los países restantes, el procedimiento seguido involucraba, generalmente, el uso de datos del último censo, a nivel del segmento censal. Primero se definieron los dominios muestrales. En Costa Rica, por ejemplo, estos fueron el Area Metropolitana de San José, el Valle Central urbano, el Valle Central rural, el resto urbano y el resto rural. En los dominios urbanos, por lo general, se seleccionaron directamente, en una sola etapa, los segmentos a entrevistar, con probabilidad proporcional a su tamaño (*PPT*), después de haber hecho una estratificación implícita. Luego se enviaron listadores al campo para preparar una lista de viviendas. Más tarde, en la oficina, se seleccionaron las viviendas a ser visitadas, aplicando un intervalo de selección proporcional a la probabilidad con que fue seleccionado el segmento.

La selección de segmentos en áreas rurales se hizo generalmente en dos etapas, por el afán de no dispersar demasiado la muestra y así ahorrar tiempo de viaje y combustible. En Costa Rica, para seguir con el mismo ejemplo, en el dominio "resto rural" se seleccionaron primeramente 26 cantones con *PPT*. Luego se seleccionaron tres segmentos censales por cantón, también con *PPT*. Después de haberse listado las viviendas en los 78 segmentos, se procedió a la selección de viviendas, usando un intervalo proporcional a la probabilidad incondicional con que fue elegido el segmento.

Una de las lecciones de la experiencia de la Encuesta Mundial de Fecundidad, y que se ha aplicado en Ecuador, es que, si los materiales censales y cartográficos a nivel del segmento son adecuados, no vale la pena utilizar más que una etapa en la selección de segmentos rurales. Esto se debe a dos consideraciones: *a*) el uso de más de una etapa disminuye la eficiencia de la muestra (un hecho ya bien conocido); y *b*) los ahorros resultantes de condensar o conglomerar la muestra son ficticios.

En el área rural de los países en vías de desarrollo, se ha gastado mucho más en el transporte *dentro* de las *UPM* (unidades primarias de muestreo) seleccionadas que en viajes *entre* las *UPM*. Esto se debe a los muy malos caminos que conectan los centros poblados pequeños dentro de una *UPM* dada, por un lado, y la existencia de caminos adecuados —o carreteras provinciales o nacionales— que sirven para viajar entre las *UPM*, por el otro. Así, en Ecuador se decidió utilizar una sola etapa en la selección de segmentos a entrevistar, tanto en áreas rurales como en áreas urbanas. Por supuesto, si se tratara de un país muy grande, sin redes de carreteras, la situación cambiaría.

Hay un último aspecto del cuadro 1 que merece ser comentado.



El número medio de entrevistas logradas por unidad última de muestreo (*UUM*) varía significativamente entre países (desde 3,2 en México hasta 29,4 en Ecuador). En tres países (México, Perú y Colombia) el número muy reducido de entrevistas por *UUM* se debe al uso de una muestra maestra ya existente. En la República Dominicana, el promedio (4,7) se debe a que la muestra para la entrevista intensiva individual fue una submuestra (de 1 en 4) de una encuesta más grande en que datos muy sumarios sobre la fecundidad fueron recogidos con una hoja de ruta. En realidad, el número de mujeres en edad fértil por *UUM* promedió 19. En los demás países, el promedio varía entre 8 y 13, con la excepción de Ecuador, donde como hemos visto, la *UUM* equivale a la *UPM*. El número de entrevistas por *UUM* tiene implicancias tanto para los errores muestrales como para los costos de recoger la información deseada, como veremos más adelante.

### 3. ERRORES MUESTRALES

Como se ha señalado, las muestras de la EMF son mensurables, o sea, el diseño muestral proporciona la base para calcular, a partir de la muestra misma, estimaciones de la variabilidad muestral. La mayor parte de los informes generales de resultados publicados por los países participantes incluyen información detallada sobre errores muestrales para una gama de variables para la muestra total, y para sub-clases de interés sustantivo. Esta labor ha sido facilitada por el desarrollo, en Londres, de un programa de computador, denominado CLUSTERS, que calcula rápida y económicamente una variedad de estadísticas concernientes a errores muestrales para un gran número de variables y subclases, en forma simultánea.

Como un ejemplo ilustrativo, en el cuadro 2 se presentan estimaciones de errores muestrales para cuatro variables de interés, de cuatro países latinoamericanos. En cada caso, se presentan resultados para la muestra total, y para una subclase socioeconómica pequeña (usualmente de menos de 500 mujeres). Las cifras presentadas corresponden a: el tamaño no ponderado de la muestra; el estimado muestral del promedio o de la proporción; el error estándar relativo (o sea, como porcentaje del promedio); y el efecto del diseño (deft). Esta última estadística es una medida de la pérdida en precisión muestral debido al hecho de no haberse utilizado muestreo aleatorio simple. Se define como la razón entre el error estándar estimado para la muestra y el error estándar estimado para una muestra aleatoria simple del mismo tamaño. A continuación se presentan algunos comentarios sobre los resultados.

ERRORES MUESTRALES DE ALGUNAS VARIABLES: PARA LA MUESTRA  
TOTAL Y PARA UNA SUBCLASE PEQUEÑA - SEGUN PAIS

MUESTRA TOTAL

	México	Colombia	Perú	Costa Rica
<b>Edad primera unión</b>				
$n$	4.166	2.141	3.807	2.064
$\bar{X}$	18,4	18,9	18,9	19,5
100.e.s./ $\bar{X}$	0,5	0,5	0,3	0,3
deft	1,76	1,35	1,05	1,02
<b>Nacidos vivos tenidos en los últimos 5 años</b>				
$n$	4.190	2.039	3.850	2.027
$\bar{X}$	1,34	1,08	1,21	0,78
100.e.s./ $\bar{X}$	2,0	3,0	1,5	2,8
deft	1,58	1,31	1,04	1,04
<b>Tamaño deseado de familia.</b>				
$n$	5.511	2.807	4.872	2.671
$\bar{X}$	4,52	4,11	3,84	4,74
100.e.s./ $\bar{X}$	1,4	1,4	1,0	1,4
deft	1,74	1,18	1,32	1,12
<b>°/o que ha usado un anticonceptivo</b>				
$n$	7.310	5.378	5.640	3.037
$\bar{X}$	39,2	37,3	48,5	81,6
100.e.s./ $\bar{X}$	3,2	4,2	2,5	1,0
deft	2,19	2,35	1,84	1,10

$n$  = Tamaño de muestra no ponderada;  $\bar{X}$  = promedio/proporción estimado; 100.e.s./ $\bar{X}$  = error estándar como porcentaje del promedio; deft = efecto del diseño.

ERRORES MUESTRALES DE ALGUNAS VARIABLES: PARA LA MUESTRA  
TOTAL Y PARA UNA SUBCLASE PEQUEÑA - SEGUN PAIS

MUESTRA PEQUEÑA

	México	Colombia	Perú	Costa Rica
<b>Edad primera unión</b>				
n	358	219	506	306
$\bar{X}$	16,8	20,0	17,9	19,7
100.e.s./ $\bar{X}$	1,2	1,1	0,6	0,9
deft	1,16	1,10	0,86	0,96
<b>Nacidos vivos tenidos en los últimos 5 años</b>				
n	316	198	422	314
$\bar{X}$	1,91	0,76	1,85	0,79
100.e.s./ $\bar{X}$	3,7	7,7	3,1	6,6
deft	1,29	0,94	1,23	0,98
<b>Tamaño deseado de familia.</b>				
n	325	304	443	404
$\bar{X}$	4,82	3,35	3,90	4,68
100.e.s./ $\bar{X}$	3,5	3,2	3,1	3,2
deft	1,06	1,14	1,24	0,95
<b>% que ha usado un anticonceptivo</b>				
n	379	354	527	452
$\bar{X}$	19,5	83,6	24,9	84,7
100.e.s./ $\bar{X}$	11,7	2,9	9,9	2,1
deft	1,12	1,22	1,31	1,03

n = Tamaño de muestra no ponderada;  $\bar{X}$  = promedio/proporción estimado; 100.e.s./ $\bar{X}$  = error estándar como porcentaje del promedio; deft = efecto del diseño.

(véase Verma, 1980).

a) Considerando las cifras del deft para las muestras totales, hay bastante variabilidad entre los cuatro países. En México, el uso de una muestra muy conglomerada (de pocas *UPM* y de muchas *USM* y *UUM*), ha resultado en un aumento en la varianza de tres veces ( $\text{deft}^2$ ) respecto a una muestra aleatoria simple. En cambio, en el Perú, donde también se utilizó un marco muestral pre-existente, pero con bastantes *UPM*, los deft para variables demográficas giran alrededor de 1,05. La muestra ad-hoc de Costa Rica, muy dispersa para un país tan chico, arroja las cifras más bajas del deft.

b) Los valores del deft difieren según el tipo de variable sustantiva: son generalmente más bajos para variables demográficas, y más altos para actitudes y variables asociadas a la anticoncepción, lo que refleja el grado de dispersión o de concentración espacial de estos atributos.

c) Altos valores del deft para la muestra total tienen poca significancia real, ya que los errores estándar son pequeños, debido al tamaño tan grande de las muestras utilizadas. El error estándar relativo gira generalmente alrededor de 1 a 3 por ciento.

d) Los errores estándares cobran mucho más importancia cuando se tratan de subclases pequeñas. En general, para subclases de tamaño 400-500, los errores relativos son del orden de 2 a 3 veces más grandes que los de la muestra total; y para subclases del tamaño 100, hasta 5 veces más grandes.

e) Según un trabajo ya citado (Verma, Scott y O'Muircheartaigh, 1980), y como se aprecia en el cuadro 2, los valores del deft son mucho *menores* en las subclases pequeñas que en la muestra total. En general, cuanto más pequeña la subclase, más se acerca a la unidad el valor del deft. Esto significa que para subclases pequeñas, la precisión muestral depende poco del diseño muestral utilizado. Se señala que estamos hablando de subclases y no de áreas geográficas.

f) Una conclusión a que han llegado los expertos en muestreo de la EMF es la siguiente: en vista de que los resultados de interés sustantivo se refieren usualmente a subclases relativamente pequeñas, y no a la población total, parecería que algunas de las muestras utilizadas en la EMF han sido demasiado ineconómicas, o sea, han sido muy dispersas, utilizando muchas *UUM* que rindieron muy pocas entrevistas. Desafortunadamente, hasta el momento ha sido imposible hacer

un estudio comparativo de los costos asociados con los diversos diseños muestrales utilizados en la EMF, por la falta de información contable suficientemente detallada. Sin embargo, actualmente se está realizando un estudio que sí proporcionará los elementos con que juzgar los distintos diseños, recogiendo datos sobre los costos del listado, los costos de entrevista, y los costos de transporte dentro y entre las UUM, las USM y las UPM.

#### 4. ERRORES NO MUESTRALES

Un componente importante del programa de la Encuesta Mundial de Fecundidad es la evaluación de la calidad de los datos. Pero, en el afán de publicar rápidamente los resultados de sus encuestas, los informes generales de los países contienen, en general, muy poca evaluación. Es así que recién ha empezado la fase de evaluación de los datos.

Hay muchas maneras de evaluar o validar datos demográficos. Su consecuencia interna puede comprobarse; su consistencia con leyes axiomáticas puede ser indagada; la preferencia por ciertos dígitos puede detectarse; y también los datos pueden ser comparados con fuentes externas (censos, estadísticas vitales, encuestas anteriores, o registros continuos de la población). Una compilación reciente de los resultados de evaluaciones de esta naturaleza hace pensar que, en general, los datos sobre fecundidad recogidos en la mayoría de los países son de una calidad sorpresivamente buena (véase Chidambaram, Cleland y Verma, 1980).

Otra técnica de evaluación es el examen de la confiabilidad de respuestas a través de una re-entrevista. Esto involucra una segunda visita a una submuestra de mujeres ya entrevistadas, después de un intervalo de varias semanas o unos meses, utilizando el mismo cuestionario o una parte del mismo. Estos programas de re-entrevistas se han llevado a cabo en diez países participantes en el programa de la EMF, y los resultados han sido publicados para tres países: Fiji, Indonesia y el Perú. En forma resumida, se presentan a continuación los principales hallazgos del proyecto de re-entrevistas en el Perú. (véase O'Muirheartaigh, Marckwardt y Verma, 1979; O'Muirheartaigh y Marckwardt, 1980).

a) Los resultados arrojados por la primera y la segunda entrevista son muy parecidos. Al examinar las primeras cuatro columnas del cuadro 3, se aprecia que las medidas de distribución —el promedio y la desviación estándar—, son básicamente idénticas para una serie de varia-

Cuadro 3

MEDIDAS DEL GRADO DE ACUERDO O DESACUERDO ENTRE  
RESPUESTAS A LA PRIMERA Y SEGUNDA ENTREVISTA, PERU

	Promedios		Desviaciones estándar		°/o de desacuerdo		
	$E_1$	$E_2$	$E_1$	$E_2$	B	D	$\alpha$
Edad actual (años)	33,7	33,7	8,7	8,8	0,06	34,1	0,98
Edad a primera unión	19,8	20,0	4,1	4,2	0,07	54,2	0,80
Año calendario del primer na- cido vivo	63,6	63,4	8,5	8,5	0,08	29,1	0,98
Intervalo proto- genésico (me- ses)	13,4	9,1	24,7	32,6	0,21	72,2	0,40
Año calendario del último na- cido vivo	72,6	72,5	4,8	4,8	0,06	22,2	0,98
Hijos nacidos en los últimos cin- co años	1,11	1,05	1,02	1,02	0,05	15,9	0,91
Total hijos naci- dos vivos	4,65	4,67	---	---	0,02	12,4	0,98
Número de hijos deseados	3,78	3,57	---	---	0,10	59,8	0,42

-- Dato no disponible.

$E_1$  y  $E_2$  se refieren a la primera y la segunda entrevista, respectivamente.

Ver el texto para las definiciones de B, D y  $\alpha$ .

bles. Por ejemplo, en el caso de la variable "edad a la primera unión", los promedios difieren en dos décimos de un año, o sea, en uno por ciento, y las desviaciones estándares en un décimo de un año. De un total de veinte variables —de todo tipo— que se han examinado, solamente dos presentan diferencias en distribución que responderían positivamente a una prueba de significancia —si fuera lícita hacerla. En la quinta columna del citado cuadro, se presenta la estadística B, una medida (entre muchas posibles) del nivel de desacuerdo entre las distribuciones marginales. 2] El valor 1 significa desacuerdo completo, y el valor 0 acuerdo completo entre las dos distribuciones marginales. Valores menores de 0,1 significan bastante acuerdo. Nuevamente se aprecia el alto grado de similitud entre los resultados de la primera y la segunda entrevista.

b) Aun existiendo gran similitud en las distribuciones, pueden encontrarse bastantes discrepancias, a nivel individual, en las respuestas a las dos entrevistas. La medida D (en la sexta columna del cuadro 3) es simplemente el porcentaje de mujeres que dieron una respuesta diferente en una y la otra entrevista. Se aprecia que, en más de un tercio de los casos, hubo discrepancias en la variable "edad actual", y en más de la mitad, discrepancias en la variable "edad a la primera unión". La variable compuesta, "intervalo protogenésico en meses", muestra más de 70 por ciento de discrepancias. Las variables compuestas, que miden el largo del intervalo entre dos acontecimientos, muestran, en general, las tasas más altas de discrepancias. También lo hacen las actitudes, como es el caso de la variable "número de hijos deseados". Esto se da sistemáticamente en todo el mundo. En el caso del Perú, se nota también que la información sobre matrimonios o uniones es mucho menos confiable que la información acerca de los nacimientos tenidos por la mujer.

c) ¿Cuáles son los factores que explican el grado de confiabilidad de respuestas? Parecería que el factor más importante es el que se refiere a las características de las entrevistadas mismas. Mujeres jóvenes, con niveles altos de educación y con residencia en áreas urbanas, dan respuestas mucho más confiables que las mujeres mayores, o que las que nunca asistieron a la escuela, o las residentes en áreas rurales. Por ejemplo, el porcentaje de discrepancias en la información sobre el mes y año del nacimiento del último hijo es de 20 por ciento para mujeres menores de 25 y de 49 por ciento para mujeres mayores de 45; es de 14 por ciento para mujeres con primaria completa, versus 58 por cien-

$$2] \quad B = \frac{2}{\pi} \cos^{-1} \left[ \sum_i (P_i P_{.i})^{1/2} \right]$$

to para mujeres sin instrucción; y es de 17 por ciento para mujeres residentes en Lima u otras ciudades grandes, versus 56 por ciento para mujeres rurales. Tomando como variable dependiente el número de discrepancias entre la primera y la segunda entrevista, se encuentra que más del 30 por ciento de la varianza puede ser explicada por las características de las entrevistadas. Dos factores más juegan un rol explicativo en el número de discrepancias: la identidad de la entrevistadora, y el tiempo transcurrido entre las fechas de la primera y la segunda entrevista. Pero estos factores son mucho menos importantes que las características de la entrevistada misma.

d) ¿Qué implicancias tienen estas discrepancias en respuestas entre la primera y la segunda entrevista para los resultados sustantivos derivados de la encuesta? En primer lugar, las discrepancias a nivel individual tienden a cancelarse, produciéndose así estimaciones para la población total muy estables. En segundo lugar, en un trabajo empírico, hemos constatado que las tabulaciones bi-variable y hasta tri-variable obtenidas de la primera entrevista y la re-entrevista son muy parecidas. También se ha experimentado con regresión múltiple, usando tres variables independientes y una dependiente. Los coeficientes 'beta' son casi idénticos para las dos entrevistas. Estos resultados reconfortan.

e) Sin embargo, se sabe que al introducir un elemento aleatorio en la medición de la relación entre dos o más variables, las medidas de relación tienden a atenuarse. Por ejemplo, la correlación entre dos variables  $x$  y  $z$  se atenúa por el factor  $1 - (\alpha_x \alpha_z)^{1/2}$ , donde,  $\alpha_x$  es la correlación entre el par de observaciones de  $x$ , y  $\alpha_z$  representa los valores de  $z$ . En la última columna del cuadro 3 se presentan los valores de  $\alpha$ . Se observa, por ejemplo, que la correlación entre el intervalo protogénico y cualquiera otra variable se reduciría por un factor de 0,37, sin tomar en consideración la confiabilidad de la segunda variable. En el *Informe General de la Encuesta de Fecundidad del Perú*, hay una tabla en que se relaciona la duración del intervalo protogénico con la edad a la primera unión. Podemos ahora deducir que el grado de relación entre estas dos variables está subestimado, probablemente en un 43 por ciento, en dicho informe.

f) Finalmente, para no dejar la impresión que el caso del Perú es excepcionalmente raro, se deja constancia que los valores de  $D$  (porcentaje de discrepancias) para la variable "edad actual" son, respectivamente, para Perú, Fiji, Indonesia y Bangladesh: 34,1 por ciento; 36,4 por ciento; 55,5 por ciento y 80,4 por ciento. Reafirmando el hecho que las discrepancias a nivel individual tienden a cancelarse, en ninguno



de estos cuatro países existe una diferencia entre la primera y la segunda entrevista, en el promedio de la "edad actual", de más de un décimo de un año.

#### BIBLIOGRAFIA

- Chidambaram, V.C., J.G. Cleland and V.K. Verma (1980), *Some Issues of Survey Methodology and Data Quality: The WFS Experience*. Paper prepared for the Population Association of America Annual Meeting, Denver. World Fertility Survey: London.
- MacDonald, A.L., P.M. Simpson and A.M. Whitfield (1978), *An Assessment of the Reliability of the Indonesia Fertility Survey Data*. *WFS Scientific Reports*, Number 3. International Statistical Institute: The Hague.
- O'Muircheartaigh, C. and A. Marckwardt (1980), *An Assessment of the Reliability of WFS Data*. Paper prepared for the World Fertility Survey Conference. World Fertility Survey: London.
- O'Muircheartaigh, C., A. Marckwardt and V. Verma (1979), *Response Reliability in WFS Data*. Paper prepared for International Statistical Institute Annual Meeting, Manila. World Fertility Survey: London.
- Scott, C. and S. Singh (1980), *Problems of Data Collection in the World Fertility Survey*. Paper prepared for the World Fertility Survey Conference. World Fertility Survey: London.
- Srikantan, K.S. (1979), *An Evaluation of the Fiji Fertility Survey Based on the Post-Enumeration Survey*. *WFS Occasional Papers*, Number 21. International Statistical Institute: The Hague.
- Verma, V. (1980), *Sampling for National Fertility Surveys*. Paper prepared for the World Fertility Survey Conference. World Fertility Survey: London.
- Verma, V. and M.C. Pearce (1978) *User's Manual for CLUSTERS*. World Fertility Survey: London.

Verma, V. C. Scott and C. O'Muircheartaigh (1980), *Sample Designs and Sampling Errors for the World Fertility Survey*. Paper prepared for the meetings of the Royal Statistical Society, London. Published in the Journal of the R.S.S. Soc. A, 143, part 3, 1980.

World Fertility Survey (1975) *Manual on Sample Design. WFS Basic Documentation*, Number 3. International Statistical Institute: The Hague.