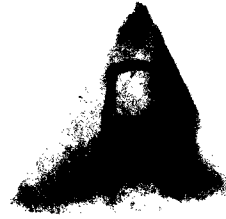


CELADE

DOCUMENTO
MICROFILMADO

CELADE - Centro Latinoamericano de Demografía
Programa de Demografía



EVALUACION DE LAS TASAS DE MORTALIDAD
OBTENIDAS CON DEFUNCIONES REGISTRADAS POR
COMPARACION CON LAS QUE SE DERIVAN DE
LA MORTALIDAD DERIVADA DE DOS CENSOS

Jorge Arévalo

Junio 1979

Documento de Trabajo (Circulación restringida)

INTRODUCCION

En este artículo se presentará un método elaborado hace un tiempo por William Brass para comparar las medidas de la mortalidad que se obtienen mediante el uso del registro de defunciones con las que se obtienen calculando las defunciones de una cohorte comparando adecuadamente la población de dos censos por grupos de edades. Como muchos de los métodos indirectos de estimación, éste se apoya en algunos supuestos que deben cumplirse para que la estimación sea válida. En la práctica los supuestos suelen no cumplirse por lo que las estimaciones pueden apartarse de los valores reales. Se dice que un método es robusto cuando a pesar de la violación de los supuestos todavía se obtienen resultados cercanos a la verdad. Se tratará aquí de someter el método a condiciones controladas en que se violarán los supuestos a fin de averiguar el grado de robustez del mismo. Para ello se ha construido una población estable cerrada la que se ha proyectado por un período de diez años manteniendo constantes los parámetros de su construcción y se calcularon las defunciones del período, de acuerdo con la ley de mortalidad utilizada. Luego se ha modificado la población inicial, la final y las defunciones del período según algunos patrones arbitrarios de omisión y mala declaración de la edad para aplicar el método y observar, en estas condiciones controladas, si es posible obtener buenas estimaciones.

EL METODO

Brass compara dos versiones de las tasas de mortalidad a partir de una edad y (se usará la misma notación que el autor) hasta la edad más elevada que una persona puede alcanzar, la que se obtiene con las defunciones registradas y otra a partir de los datos censales.

Sea

$$dy = fy \cdot dy_+^* = \delta y + dy_+^+ \quad (1)$$

donde

dy es la tasa de mortalidad exacta de la población de y años y más,

dy_+^* es la tasa calculada con las defunciones registradas,

fy es el factor por el que habría que corregir la tasa anterior,

dy_+^+ es la tasa intercensal a partir de la población censada,

δy es el valor que habría que sumar a la tasa intercensal para corregirla.

De la relación (1) se deduce inmediatamente que

$$dy_+^+ = -\delta y + fy \cdot dy_+^* \quad (2)$$

Nótese que la igualdad (2) no exige que la población sea estable, lo que de partida parece tener ventajas sobre otros métodos. Nótese también que los factores de corrección están expresados en función de la edad. Ahora bien, es posible que los errores de las defunciones registradas tiendan a ser proporcionales al número de muertes de cualquier grupo de edades, por lo que fy podría considerarse de una manera burda como constante. En cambio, los errores en la estimación de las defunciones derivadas por comparación entre las personas de una misma cohorte

enumeradas en dos censos tienden a ocurrir a causa de la falta de integridad de los censos (omisión y mala declaración de la edad) o de la migración internacional. Si hubiera proporcionalidad exacta en el error censal, esto resultaría en un error absoluto constante en las tasas de mortalidad obtenidas a partir de los censos. Esto es lo que explica que en la ecuación (1) δy sea un término que se suma a la tasa intercensal para obtener la tasa verdadera.

Si se aceptara que el error de las defunciones registradas es independiente de la edad y proporcional al número de muertes y el error de los censos también fuera independiente de la edad y proporcional al número de personas de cada edad, la relación entre las tasas intercensales y las derivadas del registro sería una línea recta en que f , δ , y la tasa verdadera pueden determinarse. De paso, el valor de δ puede interpretarse como la diferencia relativa de integridad de un censo respecto del otro. Sobre esto se volverá en otro artículo.

En el supuesto de independencia de los errores de la edad, la ecuación (2) puede escribirse:

$$dy_+^+ = -\delta + f \cdot dy_+^* \quad (3)$$

Este método puede compararse con otro presentado también por Brass hace varios años:

$$N_y/P_y = ry_+ + fy \cdot dy_+^*$$

donde

N_y es el número de personas a la edad exacta y ,

P_y es la población acumulada desde la edad exacta y hasta la edad final en la población,

ry_+ es la tasa de crecimiento de la población P_y .

También en este método, si supone la independencia de la tasa de crecimiento y del factor de corrección de la tasa de mortalidad, de la edad de las personas, se establece una relación rectilínea entre N_y/P_y y dy_+^* .

Si se hace una representación gráfica de los valores dy_+^+ y dy_+^* y de N_y/P_y y dy_+^* , se podrán observar las relaciones entre las defunciones registradas, los cambios intercensales en el número de los sobrevivientes y la distribución por edades de la población. Si los datos estuvieran totalmente exentos de error se observarían dos rectas paralelas de pendiente igual a 1 (inclinada en un ángulo de 45°), la recta correspondiente al par dy_+^+/dy_+^* pasaría por el origen y la otra interceptaría el eje de las ordenadas en un valor equivalente al de r , la tasa de crecimiento de la población de un grupo de edades cualquiera. Las desviaciones que se observan en la práctica ponen en evidencia los defectos de los datos básicos.

Brass afirma que en circunstancias favorables, las comparaciones sugerirán los supuestos apropiados sobre los que deberán basarse las correcciones.

APLICACION DEL METODO

Desde luego no hay una norma estricta en cuanto a la manera como el analista debería agrupar los datos por edad de las personas. Sin embargo Brass sugiere algunos criterios derivados de los ensayos de Hoda Rashad, como convenientes para obtener menos distorsiones en la estimación de la población expuesta al riesgo y de las defunciones intercensales. En los cuadros 1, 2 y 3 se proporcionan los datos básicos y los resultados del cálculo. Mediante ejemplos numéricos se irá ilustrando cada paso del procedimiento. En primer lugar se acumula la población censada y las defunciones registradas desde la edad más avanzada hacia las edades más jóvenes. De esta manera se obtiene la población y las defunciones de 75 años y más, 70 años y más, etc. En el Cuadro 1, donde

Cuadro 1

POBLACION ESTABLE PROYECTADA POR DIEZ AÑOS Y
DEFUNCIONES "REGISTRADAS" DEL PERIODO.

Grupo de edades	Población		Defunciones	
	Año 0	Año 10	Período 0-10	acumuladas
0-4	15 427	19 809	3 834	12 003
5-9	13 208	16 960	328	8 169
10-14	11 546	14 825	220	7 841
15-19	10 075	12 937	330	7 621
20-24	8 741	11 224	399	7 291
25-29	7 558	9 705	365	6 892
30-34	6 526	8 380	340	6 527
35-39	5 621	7 217	339	6 187
40-44	4 818	6 186	356	5 848
45-49	4 097	5 261	397	5 492
50-54	3 438	4 415	465	5 095
55-59	2 827	3 630	545	4 630
60-64	2 244	2 881	675	4 085
65-69	1 681	2 158	781	3 410
70-74	1 137	1 460	902	2 629
75-79	1 056	834	1 727	1 727
80 y más	-	522	-	-
Total	100 000	128 404	12 003	-

aparecen los datos básicos del ejemplo, figuran así en la última columna las defunciones acumuladas. En las dos primeras columnas del Cuadro 2 aparecen las poblaciones inicial y final acumuladas de abajo hacia arriba. En la tercera columna aparece el promedio de las dos poblaciones correspondientes al mismo grupo de edades. Por ejemplo: la población inicial de 5 años y más es de 84 573 y la del mismo grupo en la población final es de 108 595. El promedio es igual a 96 584. Las defunciones "intercensales" resultan de comparar las personas de la misma cohorte al inicio y al final del período. Por ejemplo, las defunciones de la cohorte que tiene al inicio 5 años y más resultan por diferencia entre 84 573 y 76 810, o sea 7 763. Estas defunciones se asignan en promedio a las personas de 10 años y más, porque corresponden a personas que durante el decenio han pasado por edades entre 5 años y más y 15 años y más.

A partir de este momento se seguirá la idea de Hoda Rashad de tomar la edad central de cada intervalo de edad en lugar de tomar la edad inicial del grupo como hasta ahora se ha hecho en el ejemplo. Para ello se procede de la siguiente manera. Para el cálculo de N_y (la población a la edad exacta y) simplemente se suman las personas del mismo grupo de edades en las poblaciones inicial y final y se asigna el resultado a la edad central del grupo. Por ejemplo la suma de la población de 5-9 al inicio y al final del intervalo, representa aproximadamente el número de personas de edad exacta 7,5 (en realidad equivale a 10 veces el número de personas de edad exacta 7,5, por ser de 10 años el intervalo intercensal). En cifras, $13\ 208 + 16\ 960 = 30\ 168$, número este último que aparece en el Cuadro 3 en la columna encabezada por $10N_y$ frente a la edad exacta de 7,5 años. El número de personas de una edad exacta y y más se obtiene promediando dos grupos consecutivos del promedio intercensal del Cuadro 2. Por ejemplo, $P_{7,5}$ resulta de promediar la población media de 5 años y más con la de 10 años y más, o sea $(96\ 584 + 81\ 500)/2 = 89\ 042$. Este número (en realidad 89 041, porque el modelo se elaboró con más dígitos) aparece en el Cuadro 3 en la columna P_y frente a la edad exacta de 7,5 años. Las defunciones registradas y las intercensales también se obtienen por promedio

Cuadro 2

POBLACION DE x AÑOS Y MAS (POBLACION ACUMULADA DESDE LAS EDADES MAS AVANZADAS HACIA LAS EDADES MAS BAJAS), PROMEDIO "INTERCENSAL" Y DEFUNCIONES "INTERCENSALES" ESTIMADAS.

Edad exacta x	Población acumulada		Promedio "intercensal"	Defunciones "intercensales" estimadas
	Año 0	Año 10		
0	100 000	128 404	114 202	-
5	84 573	108 595	96 584	8 365
10	71 365	91 635	81 500	7 763
15	59 819	76 810	68 315	7 492
20	49 744	63 873	56 809	7 170
25	41 003	52 649	46 826	6 800
30	33 445	42 944	38 195	6 439
35	26 919	34 564	30 742	6 098
40	21 298	27 347	24 323	5 758
45	16 480	21 6	18 821	5 398
50	12 383	15 900	14 142	4 995
55	8 945	11 485	10 215	4 528
60	6 118	7 855	6 987	3 971
65	3 874	4 974	4 424	3 302
70	2 193	2 816	2 505	2 518
75	1 056	1 356	1 206	1 671
80	-	522	-	-

Cuadro 3

TASAS DE MORTALIDAD Y "NATALIDAD" POR SECCIONES DE EDADES ACUMULADAS DESDE LA MAS AVANZADA HASTA UNA EDAD MENOR y (SE SIMBOLIZA CON y LA EDAD EXACTA CENTRAL DE UN GRUPO DE CINCO EDADES).

y	10 N _y	P _y	10 D*y	10 D ⁺ _y	Tasas anuales (por mil)		
					b _y	d*y	d ⁺ _y
7,5	30 168	89 041	8 005	8 065	33,9	9,0	9,1
12,5	26 371	74 906	7 731	7 628	35,2	10,3	10,2
17,5	23 012	62 561	7 457	7 332	36,8	11,9	11,7
22,5	19 965	51 816	7 092	6 986	38,5	13,7	13,5
27,5	17 263	42 509	6 710	6 620	40,6	15,8	15,6
32,5	14 906	34 467	6 357	6 268	43,2	18,4	18,2
37,5	12 838	27 531	6 018	5 928	46,6	21,9	21,5
42,5	11 004	21 571	5 670	5 578	51,0	26,3	25,9
47,5	9 358	16 480	5 294	5 196	56,8	32,1	31,5
52,5	7 853	12 178	4 863	4 761	64,5	39,9	39,1
57,5	6 457	8 600	4 358	4 249	75,1	50,7	49,4
62,5	5 125	5 705	3 748	3 636	89,8	65,7	63,7
67,5	3 839	3 464	3 020	2 910	110,8	87,2	84,0
72,5	2 597	1 855	2 178	2 095	140,0	117,4	112,9

de grupos consecutivos con el propósito de obtener una cifra que pueda referirse al centro de cada grupo de edad. Las tasas finalmente se obtienen por cociente con los valores de P_y haciendo la corrección necesaria para que sean anuales (divididas por 10) pues las defunciones y la población de cada edad exacta corresponden al período decenal.

PRUEBA DEL METODO

Para probar el método en condiciones controladas se construyó una población teórica según la siguiente relación:

$$N(x) = N \cdot b \cdot e^{-rx} \cdot p(x)$$

donde:

$N(x)$ es la población de edad exacta x ,

N es la población total,

b es la tasa de natalidad,

e es la base de los logaritmos naturales,

r es la tasa de crecimiento de la población,

x es la edad exacta

$p(x)$ es la proporción de personas que sobrevive a la edad exacta x respecto del número de personas a la edad exacta 0.

Se escogió un valor de $r = 0,025$ anual y la serie de $p(x)$ de una tabla de mortalidad de esperanza de vida al nacer de 58,97 años construida a partir del estándar general de Brass con $\beta=1$ y $\alpha=0,5$. El valor de N se hizo igual a 100 000 000 y b resultó ser igual a 35,58 por mil, por lo que la tasa de mortalidad es del orden de 10,58 por mil. Esta población se proyectó por 10 años manteniendo constantes la mortalidad y la natalidad (de hecho: $N_x^{10} = N_x^0 \cdot e^{10(0,025)}$ y aplicando la tabla de mortalidad men

cionada se obtuvieron las defunciones del período "registradas". Tanto a la población inicial como a la población final y a las defunciones se las modificó según diversos patrones de omisión y mala declaración de la edad, en un caso con independencia de la edad y en otro con selectividad según la edad.

En un caso se supuso que el grado de integridad censal era del orden de 80 por ciento y del registro de defunciones del 60 por ciento. Los patrones variables fueron:

Edad	Población		Defunciones	
	Grado de integridad por:		Grado de integridad por:	
	Enumeración	Error en la declaración de la edad	Enumeración	Error en la declaración de la edad
0-4	0,80	-	0,60	-
5-9	0,95	1,00	0,70	1,00
10-14	0,95	1,00	0,80	1,00
15-19	0,90	1,00	0,85	1,00
20-24	0,85	0,95	0,90	0,95
25-29	0,80	0,90	0,80	0,90
30-34	0,80	0,85	0,90	0,95
35-39	0,85	0,85	0,85	0,90
40-44	0,85	0,85	0,80	0,85
45-49	0,90	0,80	0,75	0,80
50-54	0,90	0,95	0,75	0,85
55-59	0,95	0,90	0,70	0,90
60-64	0,95	0,90	0,70	0,70
65-69	0,95	0,85	0,70	0,80

De 70 años en adelante se mantuvieron constantes los valores del grupo 65-69. El error en la declaración de la edad se trató de la siguiente manera: en la población se supuso que las personas del grupo de 45-49 declaraban una edad correspondiente al grupo inmediatamente inferior y así sucesivamente hasta el grupo de 5-9. De 50-54 se supuso

que declaraban su edad correspondiente al grupo siguiente y así sucesivamente hasta el final. El factor de integridad que figura en el cuadro precedente se interpreta como la proporción en que la población censada en cada grupo declaró una edad correspondiente al grupo. Por ejemplo, el grado de integridad por error en la edad de las personas del grupo de 25 a 29 años es de 0,90. Esto significa que el 90 por ciento de los que fueron censados en el grupo de 25 a 29 años declararon una edad entre 25 y 29 años y que el 10 por ciento declaró una edad entre 20 y 24 años.

Las estimaciones resultantes de los supuestos de omisión y mala declaración de la edad se combinaron en una sola. Por ejemplo: la población inicial verdadera de 25 a 29 años era de 7 558. De ella se censó el 80 por ciento, es decir 6 046. De este número, un 10 por ciento declaró tener entre 20 y 24 años, por lo que queda reducido a 5 442, pero se le agrega el 15 del 80 por ciento que fue censado en el grupo de 30 a 34 años, es decir, $6\ 526 \times 0,80 \times 0,15 = 783$. De aquí resulta que $5\ 442 - 783 = 6\ 225$ aparecen como perteneciendo al grupo de 25 a 29, tal como puede verse en la columna 3 de la población inicial.

En cuanto a las defunciones, se supone que desde el grupo de 35 a 39 años el corrimiento por mala declaración de la edad es hacia las edades inferiores (al grupo inmediato) y desde 40 a 44 años, hacia las edades avanzadas (también grupo por grupo).

La omisión en la población resulta poco más del 12 por ciento y la de las defunciones del 30 por ciento. Tanto los porcentajes de omisión independientes de la edad en la población y las defunciones, cuyo efecto se aprecia en las columnas 2 del Cuadro 4, como los patrones combinados de errores cuyo efecto aparece en las columnas 3 de dicho cuadro, parecen ser bastante severas. Así se podrá poner a prueba la robustez del método.

Cuadro 4

POBLACION ESTABLE CERRADA PROYECTADA POR DIEZ AÑOS Y DEFUNCIONES DEL PERIODO, SEGUN DIFERENTES PATRONES DE OMISION Y MALA DECLARACION DE LA EDAD.

Grupos de edades	N° Población inicial			N° Población final			Defunciones del período		
	1	2	3	1	2	3	1	2	3
0-4	15 427	12 342	12 342	19 809	15 847	15 847	3 834	2 301	2 301
5-9	13 208	10 567	12 548	16 960	13 508	16 112	328	197	229
10-14	11 546	9 236	10 968	14 825	11 860	14 084	220	132	176
15-19	10 075	8 060	9 439	12 937	10 349	12 120	330	198	298
20-24	8 741	6 993	7 663	11 224	8 979	9 840	399	239	370
25-29	7 558	6 047	6 225	9 705	7 764	7 993	365	219	278
30-34	6 526	5 221	5 154	8 380	6 704	6 618	340	204	319
35-39	5 621	4 496	4 675	7 217	5 774	6 003	339	203	259
40-44	4 818	3 854	4 218	6 186	4 949	5 416	356	214	242
45-49	4 097	3 278	2 950	5 261	4 208	3 788	397	238	281
50-54	3 438	2 751	2 940	4 415	3 532	3 775	465	279	356
55-59	2 827	2 261	2 572	3 630	2 904	3 302	545	327	396
60-64	2 244	1 795	2 187	2 881	2 305	2 808	675	405	369
65-69	1 681	1 344	1 570	2 158	1 726	2 016	781	469	579
70-74	1 137	910	1 158	1 460	1 168	1 487	902	541	614
75-79	1 056	845	1 165	834	667	882	1 727	1 036	1 335
80 y más	-	-	-	522	417	614	-	-	-
Total	100 000	80 000	87 774	128 404	102 721	112 705	12 003	7 202	8 402

1 - Integridad igual al 100 por ciento

2 - Integridad igual al 80 por ciento en la población y 60 por ciento en las defunciones.

3 - Integridad y mala declaración de la edad selectivas según la edad (Ver explicación en el texto).

Las variantes de poblaciones y defunciones del Cuadro 4 se relacionaron de diversas maneras para simular condiciones diferentes. Se hicieron en total siete ensayos que, con referencia al Cuadro 4 se enumeran:

Caso	Población inicial	Población final	Defunciones registradas
A	1	1	1
B	1	2	1
C	2	1	1
D	1	1	2
E	1	3	1
F	1	1	3
G	3	3	3

ANALISIS DE LOS RESULTADOS DE LA PRUEBA

Las tasas obtenidas aparecen en el Cuadro 5. El caso A supone que no hay errores en los datos básicos, por lo que se está en las condiciones ideales en que los valores se alinearán según una recta de pendiente igual a 1 y que pasará por el origen, cuando las tasas de mortalidad con defunciones registradas se relacionan con las intercensales (dy_+^*/dy_+^+) o que cortará la ordenada a una altura igual a la tasa de crecimiento cuando las tasas de mortalidad con defunciones registradas se relacionan las densidades de población ($dy_+^*/Ny/Py$). Por comodidad se llamará a este último método, método I, por ser el más antiguo. Al que se presenta en este artículo se le llamará método II. Puede apreciarse en el Gráfico 1 que la estimación que resulta de aplicar el criterio de agrupamiento sugerido por Hoda Rashad conduce a resultados muy precisos. Se observa lo esperado aunque parece haber una ligera desviación hacia la derecha en las tasas de las edades más avanzadas, en particular desde las de 62,5 años y más. Esta desviación podría deberse a alguna imprecisión en el

Cuadro 5

TASAS DE NATALIDAD (by_+); TASAS INTERCENSALES DE MORTALIDAD (dy_+)
Y TASAS DE MORTALIDAD REGISTRADA (dy_+^*) DE LOS SIETE CASOS

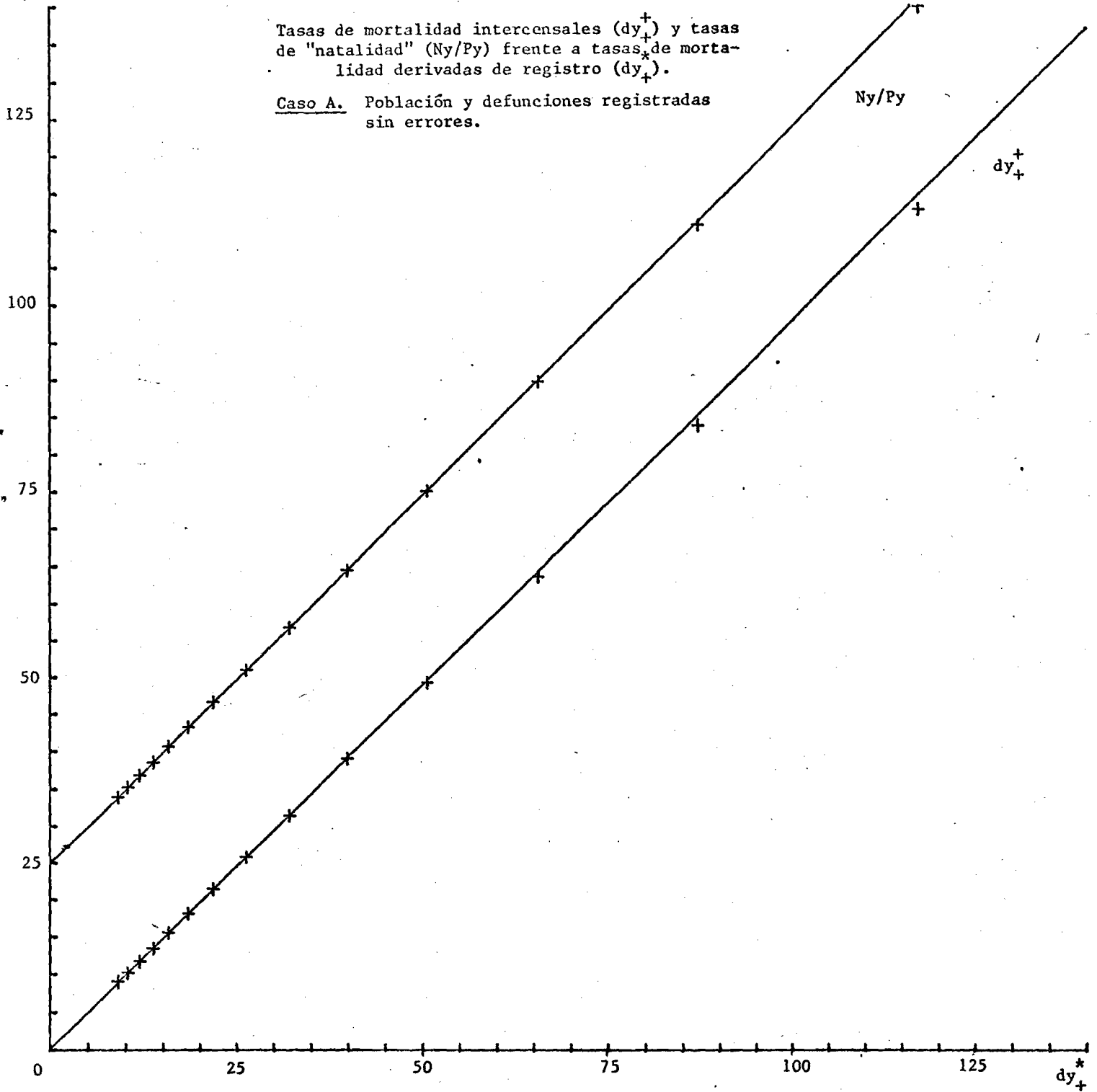
	A	B	C	D	E	F	G
by_+							
7.5	33.881	33.881	33.881	33.879	35.162	33.881	36.288
12.5	35.205	35.206	35.203	35.205	36.798	35.205	38.217
17.5	36.783	36.783	36.782	36.783	38.472	36.783	39.999
22.5	38.531	38.530	38.529	38.530	39.120	38.530	39.659
27.5	40.611	40.610	40.610	40.609	39.910	40.609	39.273
32.5	43.247	43.247	43.246	43.246	41.345	43.246	39.629
37.5	46.629	46.629	46.625	46.629	45.408	46.629	44.315
42.5	51.012	51.015	51.007	51.012	50.828	51.012	50.664
47.5	56.780	56.778	56.780	56.781	50.529	56.781	45.153
52.5	64.491	64.486	64.486	64.484	60.971	64.484	58.092
57.5	75.076	75.079	75.076	75.075	72.150	75.075	69.829
62.5	89.821	89.843	89.819	89.830	88.639	89.830	87.712
67.5	110.809	110.824	110.811	110.818	105.260	110.818	101.028
72.5	140.034	140.015	139.988	139.981	135.818	139.981	132.798
dy_+							
7.5	9.058	31.519	-12.790	9.056	22.286	9.056	10.003
12.5	10.183	32.634	-11.653	10.183	24.935	10.183	13.054
17.5	11.719	34.188	-10.136	11.718	27.682	11.718	14.433
22.5	13.482	35.974	-8.396	13.480	29.635	13.480	14.710
27.5	15.573	38.088	-6.325	15.572	30.729	15.572	14.492
32.5	18.187	40.730	-3.736	18.186	31.879	18.186	15.527
37.5	21.531	44.114	-0.428	21.531	34.384	21.531	19.225
42.5	25.857	48.498	3.848	25.858	36.408	25.858	22.716
47.5	31.530	54.255	9.446	31.530	37.761	31.530	24.490
52.5	39.100	61.956	16.891	39.098	42.041	39.098	31.222
57.5	49.406	72.492	26.988	49.408	50.160	49.408	44.211
62.5	63.729	87.227	40.934	63.740	62.119	63.740	59.265
67.5	83.993	108.254	60.467	84.001	78.482	84.001	77.391
72.5	112.922	138.679	87.921	112.896	102.407	112.896	103.075
dy_+^*							
7.5	8.990	10.129	9.853	5.394	9.600	6.723	7.58
12.5	10.321	11.628	11.311	6.192	11.100	7.722	8.82
27.5	11.919	13.428	13.061	7.150	12.924	8.866	10.29
22.5	13.687	15.420	14.999	8.211	14.930	10.060	11.81
27.5	15.785	17.783	17.298	9.469	17.219	11.501	13.50
32.5	18.445	20.780	20.213	11.065	19.996	13.318	15.45
37.5	21.859	24.626	23.953	13.114	23.507	15.624	17.85
42.5	26.207	29.616	28.806	15.771	28.161	18.779	21.30
47.5	32.121	36.189	35.199	19.271	33.922	22.993	25.31
52.5	39.932	44.988	43.757	23.957	41.102	28.502	30.01
57.5	50.674	57.086	55.524	30.399	51.296	35.985	36.71
62.5	65.093	74.014	71.984	39.411	65.751	47.544	47.61
7.5	87.475	98.219	95.516	52.291	85.471	64.617	63.01
72.5	117.438	132.301	128.647	70.422	112.733	88.506	82.41

dy_+^+ 6 Ny/Py (por mil)

Gráfico 1

Tasas de mortalidad intercensales (dy_+^+) y tasas de "natalidad" (Ny/Py) frente a tasas* de mortalidad derivadas de registro (dy_+^*).

Caso A. Población y defunciones registradas sin errores.



cálculo de las defunciones "registradas", aunque se cree que más bien se debe al método de agrupación cuando se opera en los extremos de la distribución por edades de la población en que el supuesto de variación li neal es menos sólido. Se hicieron varios ensayos de ajuste de una recta tomando diferente número de pares de tasas. Por ser el caso A el ideal en que se supone no hay errores en los datos básicos podría esperarse que los catorce puntos estuvieran en una recta. Sin embargo, no es así, aunque la desviación de los puntos correspondientes a las edades más avanzadas era relativamente pequeña, como ya se ha señalado. El ajuste con los catorce puntos obtenidos por el método II no da estimaciones muy precisas de f y de δ. Los valores son respectivamente 0,96 y 0,5 por ciento, mientras que los esperados son 1,0 y 0. En principio esto significa que a las tasas obtenidas con las defunciones registradas deberían ponderarse por 0,96 y que el segundo "censo" estaría omitido en un 0,5 por ciento respecto del primero. (La "omisión" es del segundo respecto del primero porque el signo de δ es positivo. Un signo negativo indicaría lo contrario). Al tomar sólo los nueve primeros valores se mejora la estimación: $f = 0,98$ y $\delta = 0,1$ por ciento. Con estos parámetros se trazó la recta del gráfico. Las estimaciones del método I fueron mejores: con catorce puntos $f = 0,98$ y $r = 25,18$ por mil (r debía valer 25,00 por mil). Con nueve puntos $f = 0,99$ y $r = 24,98$.

En el caso B, la población del segundo censo tiene una integridad del 80 por ciento independiente de la edad. Según el método II, tiene poca importancia cuantos pares de valores se toman para el ajuste, pues los datos se alinean muy bien. El valor de $f = 0,88$ y el de $\delta = 22,4$ por ciento. Al ser la "omisión" del segundo "censo" independiente de la edad, f y δ deben ser constantes. Puede observarse que la relación entre las tasas verdaderas (las del caso A, corregidas por 0,99) y las obtenidas en el caso B es de 0,88 en todas las edades. Por otra parte, si el valor de δ se le resta a las tasas "intercensales" se llega a valores muy próximos de las tasas verdaderas, tal como debía esperarse pues se cumple con los supuestos del método. Los resultados del método I son muy coherentes

con los del método II, pues el ajuste de la recta con todos los puntos o excluyendo los de las edades avanzadas reproduce el valor de \underline{f} , y \underline{r} oscila alrededor de 25 por mil.

En el caso C, la población del primer censo tiene una integridad del 80 por ciento independiente de la edad. Como en el caso B, poco importa el número de datos que se ajustan según una recta. Los resultados son muy coherentes con los del caso anterior, salvo el signo de $\underline{\delta}$, que ahora es negativo como debía ser, ya que la omisión es del primer censo respecto del segundo y el valor absoluto un poco más bajo, 21,6 por ciento. Acerca de la discrepancia de $\underline{\delta}$ tanto en el caso B como en el caso C respecto de la omisión del 20 por ciento implícita en el modelo, se discutirá en otro artículo, como ya se dijo más arriba.

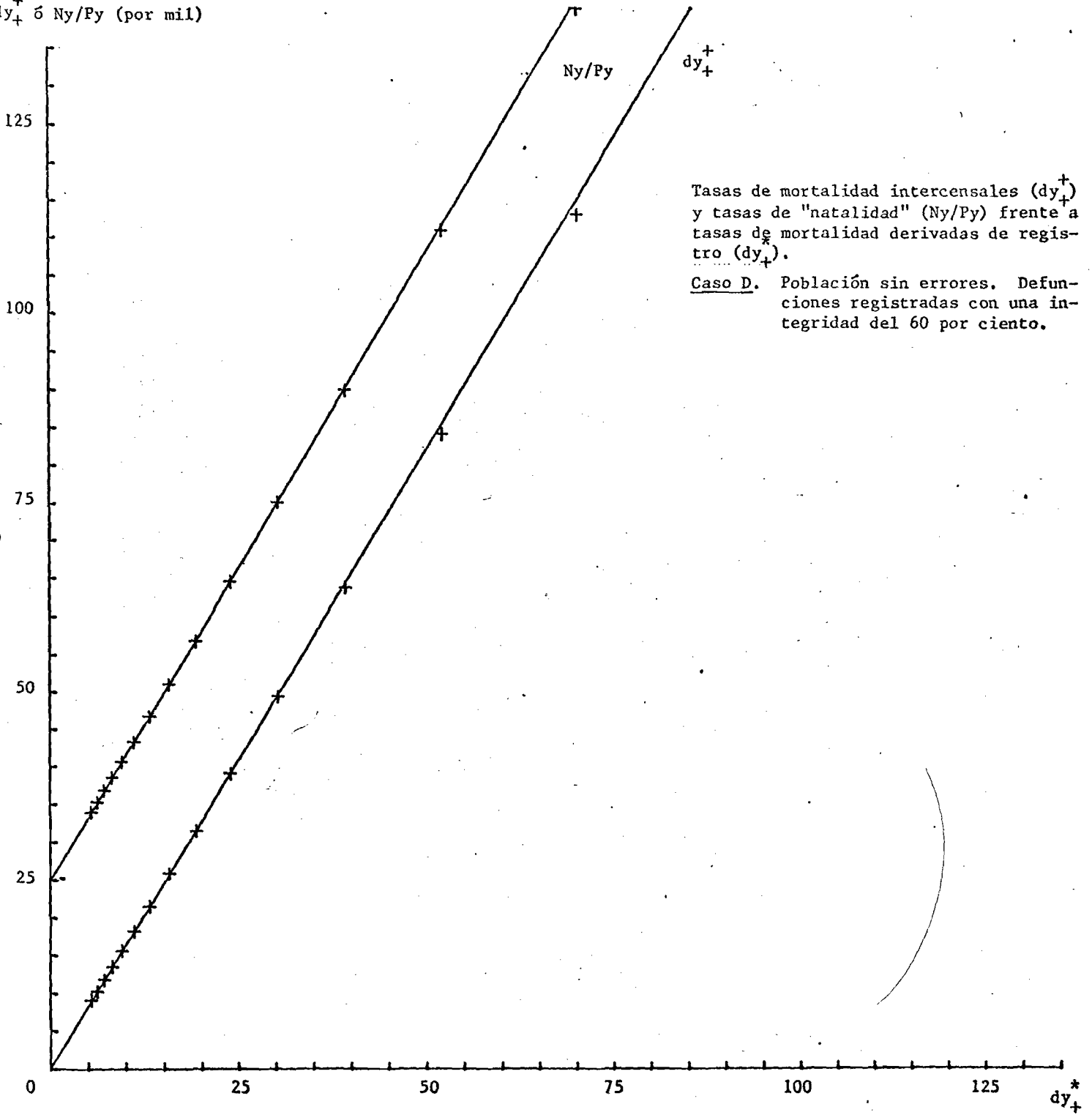
En el caso D las defunciones tienen un grado de integridad del 60 por ciento. Nuevamente los puntos se alinean, como supone el método cuando el error es independiente de la edad y las rectas pasan por el origen en el método II y corta la ordenada a la altura de la tasa de crecimiento según el método I, tal como puede observarse en el Gráfico 2. Tal como en el caso A, aquí es conveniente hacer el ajuste con los nueve primeros datos pues, aunque el valor de \underline{f} cambia poco, de 1,60 a 1,63, el de $\underline{\delta}$ pasa de 0,5 por ciento a 0,1 por ciento. Según el método I, \underline{f} pasa de 1,64 a 1,65 y \underline{r} es prácticamente 25 por mil. Las rectas se trazaron en ambos casos con los nueve primeros datos.

Los casos examinados hasta ahora, solamente tienen un valor confirmatorio de los métodos ya que no se introduce ninguna violación de sus supuestos básicos. Sirven, eso sí, para poner en evidencia la bondad del agrupamiento utilizado que, a pesar del carácter de aproximación que tiene, da resultados bastante precisos de \underline{N}_y , \underline{P}_y y \underline{D}_y^+ .

Los casos que siguen revisten más interés porque introducen violaciones más o menos severas del supuesto de independencia de los errores respecto de la edad.

Gráfico 2

dy_+^+ ó Ny/Py (por mil)



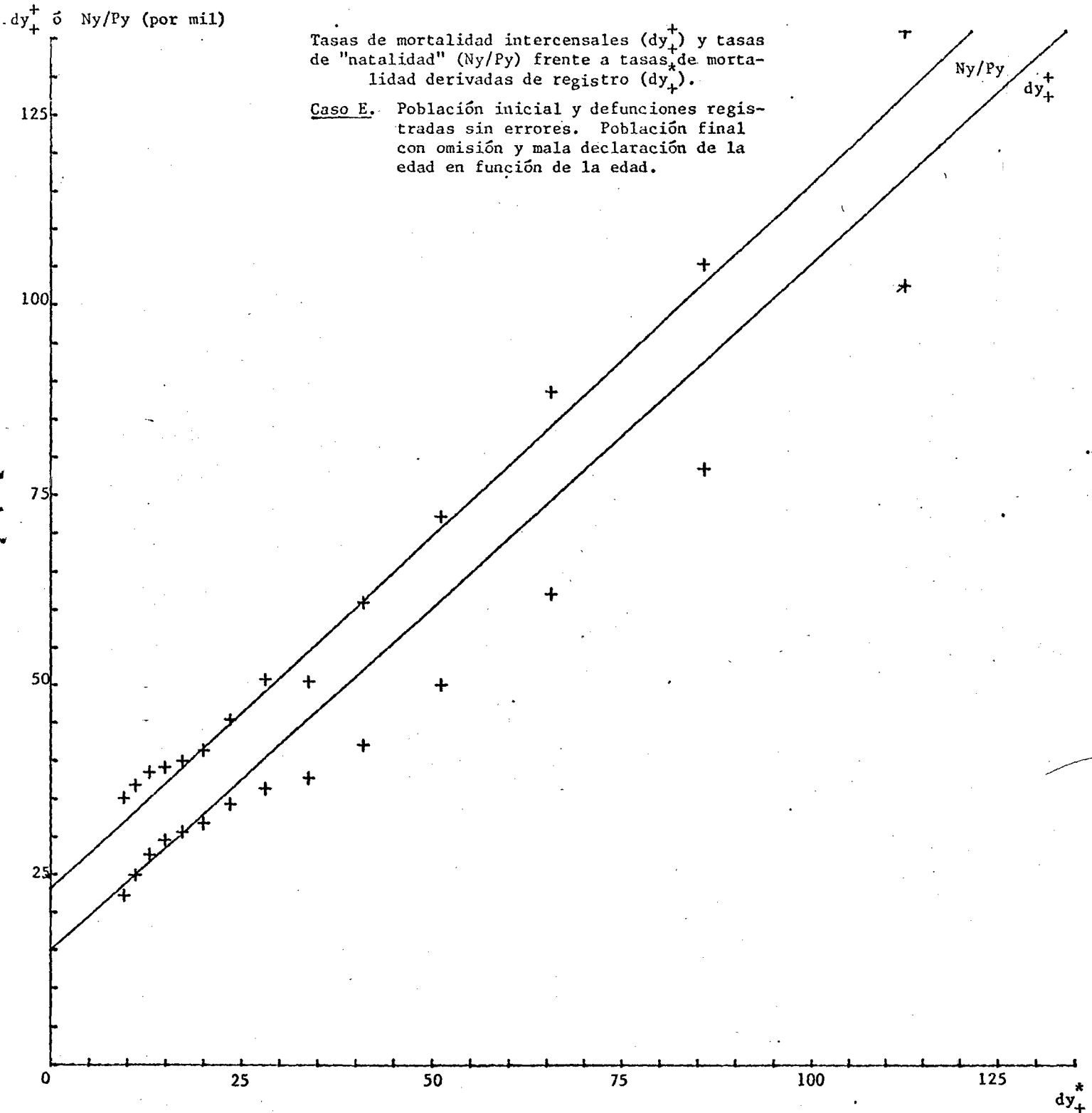
Tasas de mortalidad intercensales (dy_+^*) y tasas de "natalidad" (Ny/Py) frente a tasas de mortalidad derivadas de registro (dy_+^+).

Caso D. Población sin errores. Defunciones registradas con una integridad del 60 por ciento.

El caso E implica que el segundo censo está afectado por omisión y mala declaración de la edad selectivas por edad. Se recuerda que en el conjunto el error total es de 12,23 por ciento respecto del primer censo. Puede observarse en el Gráfico 3 la representación de los resultados de los métodos I y II, que empieza a tener la apariencia de las aplicaciones empíricas. Se tiene la impresión de que la serie representada por la relación dy_+^*/dy_+^+ es más suave que la otra, en particular si se comparan los valores del noveno par. El ajuste de los catorce puntos da un valor de $f = 0,73$ y de $\delta = 19,1$ por ciento, lo que se aparta bastante de los valores apropiados para la corrección pues el \underline{f} promedio que se obtiene de comparar las tasas del caso E con las del caso A, grupo por grupo a partir de la información del Cuadro 5 es de 0,95. Además puede agregarse que el factor de corrección de las tasas de los ocho primeros grupos varía entre 0,91 y 0,93. Se hicieron numerosos ensayos con resultados muy diversos. Por ejemplo, el ajuste de los puntos 1 a 9 dio un $f = 0,60$, muy lejos de la realidad. El de los puntos 9 a 14 dio un $f = 0,83$. Parecería, examinando atentamente los puntos (y aprovechando que se tiene la ventaja de conocer los verdaderos parámetros de la población lo que, desgraciadamente no se puede conocer en la práctica) que podría hacerse dos ajustes en lugar de uno: uno para los puntos correspondientes a las edades más avanzadas, como en el caso de 9 a 14, y otro para las edades más jóvenes. Después de varias pruebas se observó que ajustar los datos del grupo 1 al 6 daba las mejores estimaciones de \underline{f} (0,90) y de $\underline{\delta}$ (15 por ciento) con cuyos parámetros se trazó la recta que figura en el gráfico.

El ajuste de una recta a los datos del método I también tienen un campo de variación extenso, pero el hecho de conocer la tasa de crecimiento, que es un dato que puede conocerse externamente en los casos reales, ayuda a acotar los ajustes, aunque no es una ayuda muy precisa. Por ejemplo, el ajuste tomando los pares de observaciones desde el quinto al décimo $r = 25,32$ y $f = 0,84$ (algo bajo). En cambio, el ajuste desde el

Gráfico 3



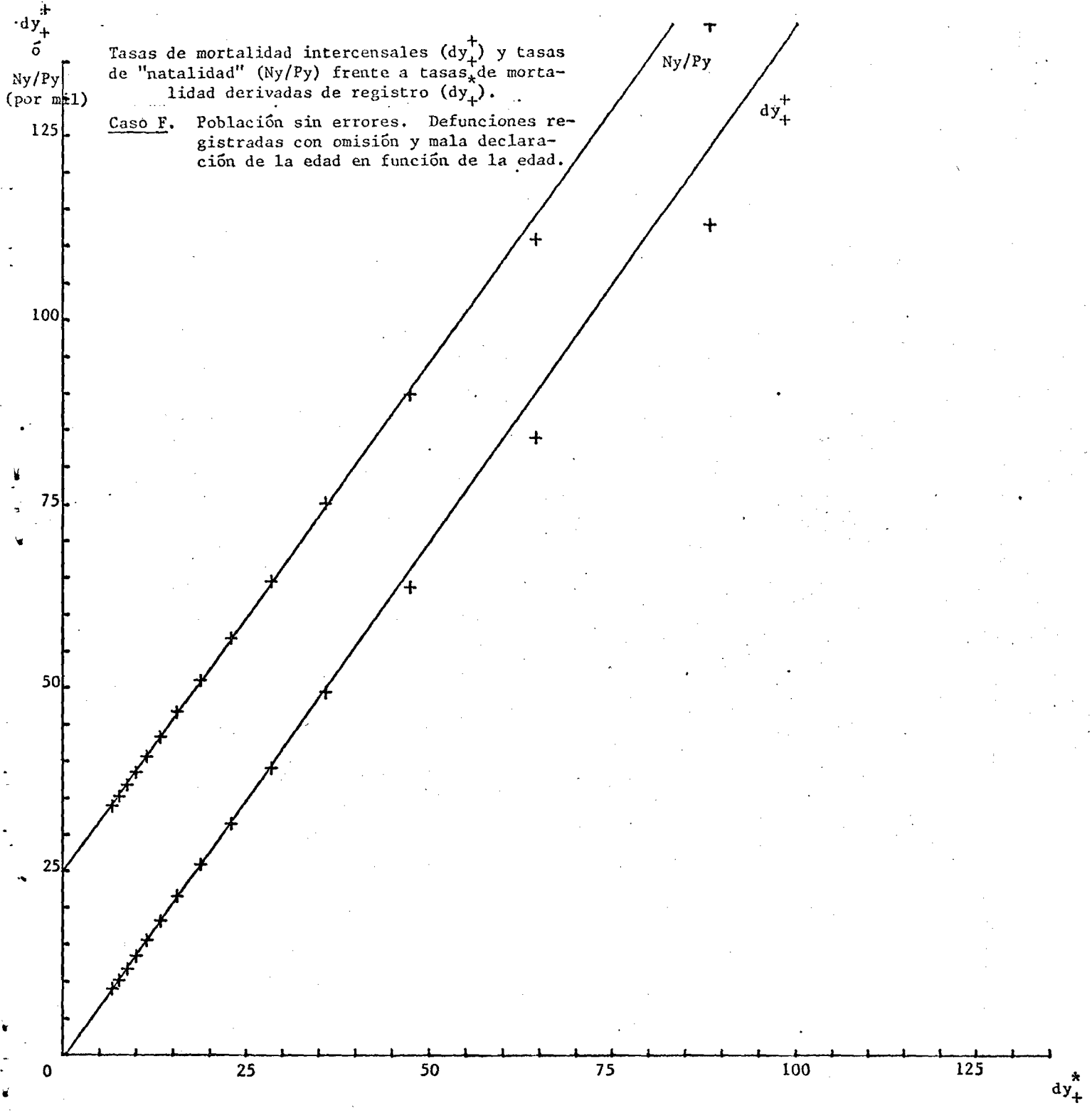
quinto al undécimo par, da un $f = 0,93$ (casi igual al promedio) pero $r = 23,1$ por mil, más lejos del 25 por mil con que se construyó la población.

Ante este caso cabe preguntarse si diferentes investigadores llegarían a la misma solución o no.

En el caso F se supone que las defunciones registradas están afectadas por omisión y mala declaración por edades selectivas por edad, de acuerdo con un patrón de errores diferente del de la población. El efecto global es de un 30 por ciento de omisión (y de 25 por ciento si se considera a los de 5 años y más) variable por supuesto en cada grupo de edades. A partir del grupo de 5 a 9 años en adelante (ver Cuadro 4) la omisión por cada grupo sucesivo es de 0,40; 0,30; 0,20; 0,10; 0,07; 0,24; 0,06; 0,24; 0,32; 0,29; 0,23; 0,27; 0,45; 0,26; 0,32 y 0,23 por unidad.

En el Gráfico 4 se representan los resultados de la aplicación de los dos métodos. A pesar de los diferentes grados de error entre los distintos grupos de edades, sorprendentemente (por lo menos a primera vista. El efecto suavizador de la acumulación es evidente) los puntos se alinean muy bien y cuesta ver en el gráfico las pequeñas fluctuaciones que hay en los primeros nueve o diez puntos. En las edades más avanzadas hay un desplazamiento hacia la derecha, pero debe recordarse que esto último también ocurre en el caso A, donde se supone que no hay errores en los datos básicos. Si se comparan las tasas dy_+^* (obtenidas con las defunciones registradas) del caso F que se está analizando con las del caso A, (ver Cuadro 5) se verá que el factor de corrección tomando cada grupo de edades por separado fluctúa entre 1,31 y 1,42, campo de variación relativamente pequeño. El promedio simple resulta ser de 1,36. Los resultados del ajuste de una recta a los datos del método II muestran un valor de f que fluctúa entre 1,28 cuando se toman los catorce datos y de 1,40 cuando sólo se toman los primeros nueve, valores ambos muy cercanos al mencionado promedio de 1,36. El valor de δ mejora cuando se excluyen los últimos cinco puntos pasando de 1,3 por ciento a -0,5 por ciento. El valor ideal

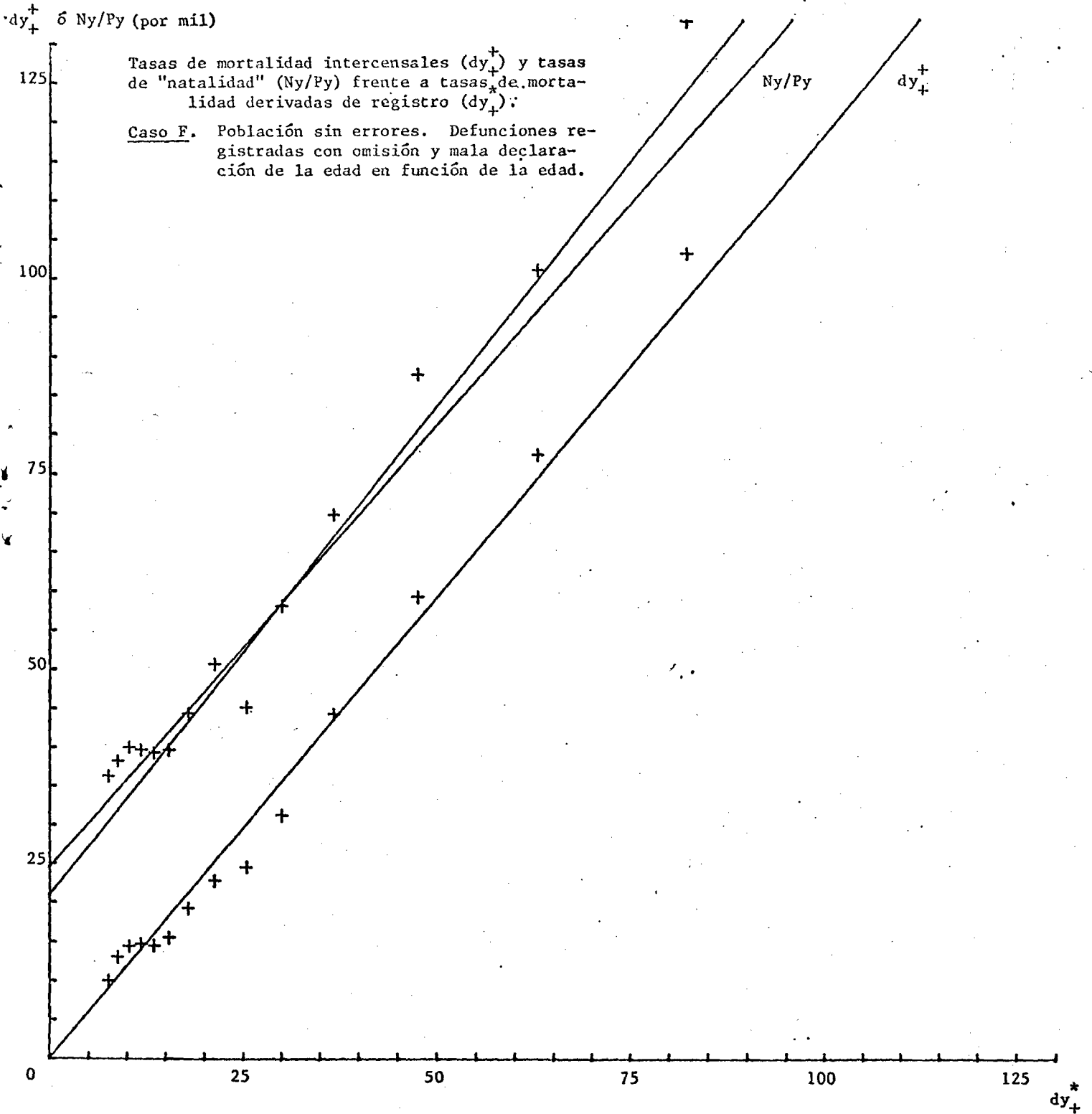
Gráfico 4



de δ debía ser 0 pues los censos no tienen errores. Sin embargo hay que recordar que el método aproximado de estimación genera su propio error, como se vió en el caso A, en que tampoco se obtuvo el valor ideal. El ajuste de la recta a los resultados del método I dan un valor de f que varía en un campo similar al del método II. Se hicieron múltiples ensayos tomando más o menos puntos y excluyendo algunos iniciales o finales y el valor varió entre 1,31 y 1,42. El valor de r osciló alrededor de 25 por mil.

El último caso, G, supone que las poblaciones inicial y final y las defunciones están afectadas por omisión y mala declaración de la edad. En las dos poblaciones se aplicó el mismo patrón de errores. Sin embargo, el cálculo de las defunciones intercensales resulta afectado pues resultan de comparar personas de dos grupos de edades diferentes, que están afectados también de manera diferente por el error. Los resultados, que aparecen en el Cuadro 5 se representaron en el Gráfico 5. Como en el caso E, la distribución según el método II parece más suave. A pesar de la irregularidad de los puntos (los primeros diez valores) parece haber una clara tendencia rectilínea de las observaciones. Sin embargo, las opciones para el ajuste no aparecen demasiado claras. El examen de los primeros nueve o diez valores parecerían pertenecer a una recta de pendiente diferente a la que sugieren los de once a catorce. Mientras no se tenga mucha experiencia en la aplicación del método las dudas inducirán a soluciones probablemente divergentes. Tal vez el uso repetido y el análisis den al analista recursos para decidir con razonable seguridad. Para ilustrar lo que se comenta basta considerar que el ajuste de los catorce puntos da un valor de $f = 1,25$ (exactamente el promedio de los factores correspondientes a las tasas individuales). En cambio si se toman los primeros nueve datos $f = 0,77$. La recta se trazó con los parámetros del ajuste de los catorce puntos.

Gráfico 5



El ajuste de los resultados del método I presentó más problemas. Los catorce puntos dan un valor de $f = 1,29$, lo que parece bastante apropiado; pero el valor de $r = 22,7$ por mil, discrepante del 25 por mil verdadero. Si se hacen otros ensayos para lograr que el valor de r sea más parecido al de la población, el valor de f se aleja del promedio. Por ejemplo, el ajuste del par 3 hasta el 11 da un $r = 25,1$ por mil, mientras que $f = 1,10$. En el Gráfico 2 se han trazado dos rectas de ajuste para mostrar la discrepancia entre un ajuste en que se ha tratado de mantener el valor de la tasa de crecimiento (si se conoce por otras fuentes, desde luego) y otro que se aproxima más al factor de corrección que debería usarse (pero desconocido en la práctica). La recta de pendiente menos marcada se trazó con los parámetros del ajuste de los pares del 4 al 10, excluido el 9, y da $f = 1,13$ y $r = 24,6$. La otra corresponde al ajuste de los pares 5 a 10, excluido el 9 que da $f = 1,21$ y $r = 22,7$.

CONCLUSIONES

Lo primero que uno debiera preguntarse es si el método II presenta ventajas sobre el método I, más antiguo. En cuanto a los datos requeridos la ventaja es para el método I pues puede aplicarse aún con datos de un solo censo. En todos los casos, excepto el último pudo apreciarse que las estimaciones de f eran más precisas o estaban más estrechamente acotadas cuando provenían del método I. Sin embargo, el último caso considerado en este ejercicio, que es el que más real parece pues toda la información básica esta afectada por errores, muestra que el ajuste ha sido relativamente más fácil cuando se trató de los resultados del método II. El examen del Gráfico 5 permite apreciar además que la secuencia de puntos se distribuye con más regularidad cuando pertenece al método II. Hay que hacer notar que la regularidad de la secuencia permite seleccionar los puntos con más seguridad. Obsérvese el noveno punto en la distribución de los resultados del método I y se apreciará que plantea serias dificultades decidir su inclusión o exclusión en un ajuste.

Los casos planteados son los primeros y más fáciles que se pueden proponer, pero no los más frecuentes. Hace falta proseguir el desarrollo en el caso de una población no estable y también en el caso de una población abierta y la combinación de ambas. Parece difícil entonces pretender generalizar algunas conclusiones que pueden derivarse de este intento. Sin embargo se han observado algunos comportamientos que no necesariamente deberían estar excluidos en ulteriores consideraciones. Se ha notado a través del análisis de los resultados y de los gráficos, que los errores que producen más alteraciones en la estimación de las tasas se derivan del censo de población y no, como podría suponerse a priori, de los errores de los registros, cuando de la violación del supuesto de independencia de los errores se trata. (Ya se vió que cuando este supuesto no se viola los resultados son los esperados). Esta observación alienta la sugerencia de intentar corregir o suavizar la distribución por edad de la población antes de aplicar el método II, con la esperanza de suavizar también la irregularidad en la distribución de las tasas.

Por último y no lo menos importante, debe subrayarse la robustez de los métodos, pues aún en los casos de violación del supuesto de independencia de los errores respecto de la edad, los valores que podían obtenerse de \underline{f} , $\underline{\delta}$ y \underline{r} eran muy semejantes a los verdaderos.