The Effect of Alternative Matching Procedures on Fertility Estimates Based on the Own-Children Method

by Michael J. Levin and Robert D. Retherford

In the own-children method of fertility estimation, enumerated children in a census or household survey are first matched to mothers within households, usually on the basis of answers to questions on relation to head of household, age, sex, marital status, and number of living children. The matching algorithm checks that relationships and ages of mother and child are compatible, that the mother is not single, and that she is not matched to more children than she states are still living. These matched (i.e., own) children, classified by own age and mother's age, are reverse-survived to estimate births by age of mother in previous years, and unmatched (non-own) children are distributed by age of mother according to the distribution of own children by age of mother. Reverse survival is similarly used to estimate the number of women by age in previous years. After adjustments are made for incorrect enumeration, age-specific birth rates are calculated by dividing the births by the women. (For methodological details, see Grabill and Cho 1965; Cho et al. 1970; Cho 1973; Ho 1978; Retherford and Cho 1978.)

The matching algorithm, which is one of several computer programs used in applying the own-children method, sometimes mismatches. For example, an adopted child may be erroneously matched instead of classified as non-own if the census or survey fails to distinguish adopted children from biological children. Or there may be more than one eligible woman in the household (same relation to head, same marital status, etc.), in which case the child may be matched to the wrong woman since the algorithm matches the child to the first eligible woman encountered in the household search. Because of the possibility of mismatch, some censuses and household surveys instruct interviewers to record mother's person number (i.e., her line number in the household listing) for each child whose biological mother is in the same household. Presumably, the direct match based on mother's person number (MPN matching) matches fewer children than the indirect match based on relation to head of household (RHH matching), since there is, in the former case, less ambiguity about the mother-child relationship and therefore a reduced probability of erroneously matching the child when the biological mother is actually absent.

Although a direct question on mother's person number increases the accuracy of matching, it also adds to the cost of the census or household survey. It is therefore of interest to examine how much difference MPN matching, compared with less costly RHH matching, makes in the ultimate fertility estimates. To this end, we examine in this paper two sets of own-children fertility estimates, one for American Samoa and the other for East Java Province in Indonesia (the choice of one province instead of all of Indonesia was made in order to economize on computing costs). In American Samoa, adoption is common, households are generally large and complex, and the frequency of in- and out-migration (mainly to the United States) is very high, leading to frequent temporary separations of family members; under these conditions, the likelihood of mismatch is much higher for RHH matching than for MPN matching. In East Java, on the other hand, households are simpler and smaller and the population is less migratory; under these conditions, the likelihood of mismatch is low for both types of matching. Our expectation is, then, that MPN matching should improve the accuracy of the fertility estimates more for American Samoa than for East Java.

Michael J. Levin is a social science analyst in the Population Division of the U.S. Bureau of the Census, and Robert D. Retherford is a research associate at the East-West Population Institute.

Data

Results for American Samoa are based on the 100 percent 1974 census count comprising 29,190 individuals. Results for Indonesia are based on the 1976 Indonesian Intercensal Population Survey, also known as SUPAS II, which included a sample of 35,822 individuals from the province of East Java. Mean household size from these data is 6.9 in American Samoa and 4.6 in East Java. Both of these counts asked mother's person number as well as relation to head of household and therefore allow computation of fertility estimates based alternatively on RHH and MPN matching.

Relation-to-head-of-household codes used in RHH matching are

for American Samoa:

1. head
2. wife of head
3. son, daughter, nephew, or niece of head
4. grandchild of head
5. brother, sister, brother-in-law, or sister-in-law of head
6. father, mother, father-in-law, or mother-in-law of head
7. other relative
8. no relation
9. unknown

for Indonesia:

1. head
2. wife of head
3. own child of head
4. non-own child of head (adopted child or stepchild)
5. grandchild of head
6. parent of head
7. parent of wife of head
8. daughter-in-law or son-in-law of head
9. other family
10. other non-family
11. unknown

Since the Indonesia codes are somewhat more detailed than the American Samoa codes, they allow a more precise match of children to mothers by means of RHH matching; hence the results reported below for the two countries are not strictly comparable. The RHH match algorithm is modified in each application to take advantage of whatever codes are available.

Findings

Table 1 shows the percentage that non-own (i.e., not-matched) children are of all children in each single-year age group according to type of match. As expected, the percentage not matched is higher in American Samoa than in East Java. Also expected is the greater divergence between RHH and MPN matching in American Samoa than in East Java. The percentage not

Table 1 Non-own children as percentage of all children by type of matching

Age	American Samoa 1974		East Java 1976	
	RHH	MPN	RHH	MPN
0	24.8	13.2	2.1	1.2
1	23.2	15.5	3.8	1.5
2	26.2	18.6	4.2	3.1
3	22.4	17.2	6.7	5.2
4	24.6	19.2	6.2	4.4
5	25.0	21.0	6.0	4.5
6	22.2	17.5	7.9	6.5
7	24.1	20.3	9.2	7.1
8	23.5	20.1	10.9	9.0
9	22.8	20.2	8.7	6.5
10	24.5	21.9	12.5	9.5
11	27.0	24.1	13.3	9.3
12	22.4	21.5	15.4	12.6
13	29.0	27.9	15.4	12.4
14	29.5	27.7	17.5	15.6

NOTES: Non-own children are those not matched to mothers within the same household. RHH denotes matching based on relation to head of household. MPN denotes matching based on mother's person number.

matched rises steeply with age of child mainly because older children are more likely than younger children to live in a household other than their mother's. An unanticipated exception is RHH matching in American Samoa, where the percentage not matched is unusually high among younger children and the rise with age in the percentage not matched is largely eliminated. The difference between RHH and MPN matching in American Samoa is especially great for younger children.

The most startling finding in Table 1 is that in both American Samoa and Indonesia the percentage not matched is higher for RHH matching than for MPN matching, the reverse of what was anticipated. To discover why the MPN match matches more children than the RHH match, we examined in detail each individual match in a household listing from a 25 percent systematic sample of American Samoa's enumeration districts. (This was not done for East Java, since discrepancies are so small.) The MPN match produced 815 unmatched, or non-own, children. Of these, only 11 were erroneously matched by the RHH match, indicating that overmatching by the RHH algorithm is not a serious problem. Offsetting these 11 children who were overmatched were 107 children who were correctly matched by MPN matching but erroneously not matched by RHH matching, owing to errors in relationship codes. These errors appear to consist mainly of respondent errors, not interviewer errors. Quite commonly, for example, the household head incorrectly reported a child as grandchild, while correctly reporting the child's mother (identified by MPN matching) as "other relative" instead of daughter or daughter-in-law as would be necessary if the child were truly a grandchild. If the household lacked an eligible daughter or daughter-in-law, RHH matching then incorrectly designated the child as non-own. This kind of error was especially common for very young children, for whom the reported grandparent-grandchild relationship evidently often reflects an affectionate social tie rather than a biological tie, and it is consistent with the large discrepancies between RHH and MPN matching at very young ages in the percentage that non-own children are of all children in the left half of Table 1. In all, because of errors in relationship codes, RHH matching produced 815 + 107 − 11 = 911 non-own children in the sample, 12 percent more than MPN matching.

The results in Table 1 suggest that fertility estimates based alternatively on RHH and MPN matching should be similar for East Java but somewhat divergent for American Samoa, particularly in the years just previous to enumeration, for which fertility estimates are based on reverse survival of very young children. Table 2 and Figures 1 and 2 confirm this expectation. In the case of East Java, fertility estimates based alternatively on RHH and MPN matching differ by less than 2 percent, and usually less than 1 percent, over single calendar years between 1962 and 1976. In American Samoa, on the other hand, the discrepancy, though small in years close to 1960, increases for years closer to the census date, reaching almost 4 percent in 1974. In American Samoa, estimates of the total fertility rate

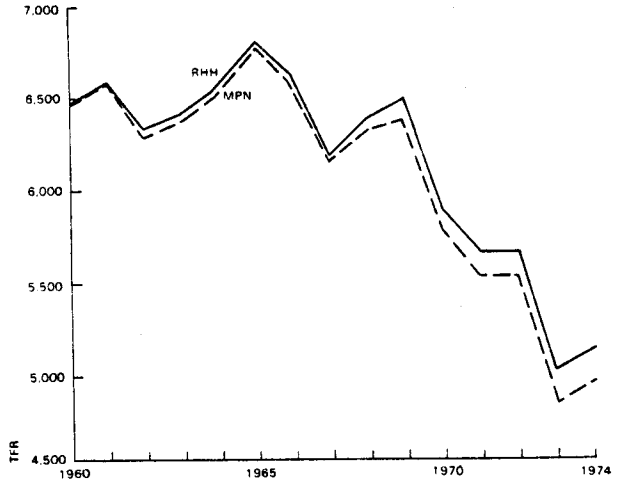

Figure 1 Own-Children Estimates of Total Fertility Rates for American Samoa. (SOURCE: Table 2.)

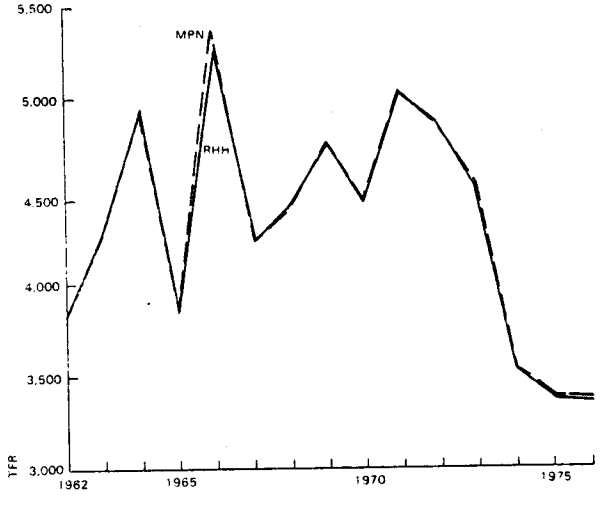

Figure 2 Own-Children Estimates of Total Fertility Rates for East Java. (SOURCE: Table 2.)

Table 2 Own-children estimates of total fertility rates according to type of matching (rates per thousand)

Years before enumeration	American Samoa				East Java			
		TFR				TFR		
	Year	RHH	MPN	RHH/MPN	Year	RHH	MPN	RHH/MPN
14	1960	6479	6469	1.002	1962	3820	—	—
13	1961	6596	6572	1.004	1963	4246	4252	.999
12	1962	6324	6291	1.005	1964	4935	4894	1.008
11	1963	6412	6373	1.006	1965	3918	3853	1.017
10	1964	6558	6532	1.004	1966	5376	5348	1.005
9	1965	6812	6769	1.006	1967	4228	4238	.988
8	1966	6628	6585	1.007	1968	4407	4391	1.004
7	1967	6207	6157	1.008	1969	4755	4749	1.001
6	1968	6397	6334	1.010	1970	4437	4445	.998
5	1969	6516	6395	1.019	1971	5016	5032	.997
4	1970	5899	5788	1.019	1972	4873	4872	1.000
3	1971	5680	5535	1.026	1973	4516	4533	.996
2	1972	5671	5532	1.025	1974	3532	3541	.997
1	1973	5028	4844	1.038	1975	3386	3401	.996
0	1974	5154	4961	1.039	1976	3359	3383	.993

NOTE: See note to Table 1.

based on RHH matching are consistently higher than estimates based on MPN matching. In East Java, the much smaller discrepancies are in both directions and show no clear pattern. The consistent direction of the discrepancy in American Samoa stems from a systematic distortion in the age pattern of fertility, to which we now turn.

Table 3 examines the effects of alternative matching procedures on the age pattern of fertility. Calendar years are grouped into two 5-year periods in order to minimize effects of heaping of children's ages on preferred digits on the fertility estimates. The effects of such heaping are evident in the jagged fertility trends observed earlier in Figures 1 and 2.

In East Java, the difference between the age patterns of fertility estimated alternatively using RHH and MPN matching is inconsequential, consistent with the small differences in matching results and TFR estimates shown earlier in Tables 1 and 2 and the figures. In American Samoa, however, RHH matching results in underestimates of fertility at the younger reproductive ages and overestimates at the older reproductive ages, especially for the second 5-year period, 1968–72. This evidently occurs because RHH matching erroneously allocates too many children to older women and not enough to younger women, relative to results based on MPN matching. The underlying reasons for this misallocation are not entirely clear, but one source of error is the method's assumption that non-own children are distributed by age of mother in the same proportions as own children. This assumption is violated in American Samoa, where non-own children are in reality concentrated among younger mothers. Because the number of women in each 5-year age group decreases with age, the shift of children

(continued on page 17)

Table 3 Own-children estimates of age-specific birth rates and total fertility rates according to type of matching (rates per thousand)

Country, period, and type of matching	15–19	20–24	25–29	30–34	35–39	40–44	45–49	TFR
American Samoa								
1963–67								
RHH	43	233	321	312	241	123	29	6519
MPN	47	242	320	311	232	117	25	6470
RHH/MPN	.923	.964	1.003	1.005	1.037	1.053	1.178	1.007
1968–72								
RHH	43	222	287	270	230	110	39	6000
MPN	47	239	293	260	215	98	27	5995
RHH/MPN	.925	.927	.979	1.037	1.067	1.126	1.435	1.001
East Java								
1965–69								
RHH	135	226	206	168	108	48	14	4532
MPN	134	225	205	167	107	48	15	4511
RHH/MPN	1.007	1.005	1.005	1.010	1.004	.996	.953	1.005
1970–74								
RHH	124	230	207	162	100	51	14	4435
MPN	125	231	206	162	101	51	15	4448
RHH/MPN	.989	.995	1.003	1.004	.997	1.007	.925	.997

NOTE: TFRs and ratios are computed from more exact values than those shown. See also the note to Table 1.

ALTERNATIVE MATCHING *(continued)*

from younger women to older women increases birth rates at the older reproductive ages more than it decreases birth rates at the younger reproductive ages. This may explain why the TFR computed from age-specific birth rates based on RHH matching slightly but consistently exceeds the TFR computed from birth rates based on MPN matching.

Proportional errors in age-specific birth rates for American Samoa are as high as 8 percent at ages 15–19 and as high as 44 percent at ages 44–49. Absolute errors in these rates are, of course, much smaller than the percentage errors suggest, because fertility in these extreme age groups is very low. Moreover, the errors are largely offsetting. Hence errors in the TFR are quite small. Note that errors in the TFR are smaller in Table 3 than in Table 2 and the figures; aggregation of calendar years reduces the discrepancy in the TFR between the two types of matching.

Conciusion

The above findings suggest that the extra coding costs necessitated by using the mother's-person-number match in place of the relation-to-head-of-household match are justifiable when migration rates are high, households are large and complex, and relationship codes are not very detailed. In this case, use of MPN matching instead of RHH matching can result in significant improvement of the fertility estimates. But when migration rates are low, household structure is simple or only moderately complex, and relationship codes are fairly detailed, RHH matching works almost as well as MPN matching, and the ultimate fertility estimates are affected very little by the choice of matching procedure. Were RHH matching expensive, MPN matching might still be advantageous. But RHH matching is very inexpensive: the cost of RHH matching at the current University of Hawaii computer charge of about US$225 per hour is approximately US$0.54 per thousand households, plus about two weeks of programming time to modify the RHH computer algorithm to fit the country's particular data format. MPN matching may, of course, be justified for reasons other than fertility estimation, as, for example, in analyses of family and household structure. In such cases, findings presented in this paper suggest that edit programs should be developed to check for consistency of responses to the relation-to-head question. Census or survey pretests can be used to ascertain whether coding of mother's person number is cost-effective. □

ACKNOWLEDGMENTS

We are grateful to the Indonesian Central Bureau of Statistics and to the Government of American Samoa for permission to use census data for this analysis. This research was supported in part by the Office of Population of the United States Agency for International Development.

REFERENCES

Cho, L. J. 1973. The own-children approach to fertility estimation: an elaboration. In International Union for the Scientific Study of Population, *International Population Conference, Liège 1973*, vol. 2, pp. 263–278.

Cho, L. J.; Grabill, W. H.; and Bogue, D. J. 1970. *Differential Current Fertility in the United States.* Chicago: University of Chicago Press.

Grabill, W. H., and Cho, L. J. 1965. Methodology for measurement of current fertility from population data on young children. *Demography* 2:50–73.

Ho, V. 1978. MATCH: a program to match children and spouses to women, stage 1 of own-children package. Honolulu: East-West Population Institute. Mimeo.

Retherford, R. D., and Cho, L. J. 1978. Age-parity-specific birth rates and birth probabilities from census or survey data on own children. *Population Studies* 32:567–581.

CENSUS CONFERENCE *(continued)*

uses of census data often cannot be anticipated. He pointed out the need for flexibility in data availability and management to make data collected in one period usable in the future.

Sampling

Answers to some of the questions about how to collect the additional data wanted by policy-makers surfaced in the session on sampling. Ironically, some of the participants shifted their stances and asked why the census should be conducted if the full-count data were not going to be used. One of the papers presented gave an account of the use of sampling in conjunction with India's 1980 census, and the other offered suggestions on the use of sampling for extensive tabulations and complex analysis.

P. Padmanabha described India's use of sampling in conjunction with the recent census. Twenty percent of the enumeration districts were selected for sampling. These districts were asked questions on migration and fertility that were not included on the full-count questionnaire. Use of the enumeration district as the areal unit was chosen in part to avoid training all enumerators in the more complex task of conducting the sample survey as would have been necessary with a sample based on a random selection of households. At the tabulation stage, an additional 20 percent of the enumeration districts was selected to make a combined sample of 40 percent for the presentation of economic data (which had been collected from all households). One quarter of the 20 percent fertility and migration sample was selected for early tabulation and presentation of national-level estimates.

Jay-Soo Park told the conference that savings from using samples for tabulation permit more tabulations and more complex tabulations. He argued that many statistics derived from census data need not be calculated from the total census count. The only statistics that must come from the full count are those that require much detail either at the small-area level or by fine categories. Full-count data can be used to check sample statistics by comparing variables available in both data sets that are likely to be correlated with the desired statistics. Samples can be selected from the data collected in the census or additional data can be collected during the census from selected households.

Park advocated using special questionnaires at randomly selected households to collect additional data during the cen-