ANNEXES

ANNEX I

IMPUTATION OF MARITAL STATUS OF HOUSEHOLD MEMBERS IN DHS SURVEYS

Many of the household schedules of the Demographic and Health Surveys (DHS) employed in the present publication did not obtain information about marital status. For the purposes of the present study, this is a significant omission, because marital status is an important determinant of living arrangements, and marital status is also strongly associated with other demographic characteristics that are of interest in this study—age and gender, particularly. However, it is often possible to infer marital status from other information available in the surveys. The present annex describes the process of imputation that was employed in the multivariate analyses reported in chapter III.

The protocol followed by DHS requires an enumeration of households, and provides a representative sample of individuals eligible for the main DHS survey as well as of households and their members. The household sample can, in many cases, substitute for a sample of census records, albeit with a significantly smaller sample size.

The information about household members includes, among other items, age, sex, relation to the head of household and, at least in the case of eligible persons, marital status. However, in many of the DHS surveys, marital status was not elicited for residents of the household other than those eligible for the main survey, namely, women of reproductive age. This lack of information on marital status poses serious problems for the study of themes related to household organization, such as the living arrangements of older persons. A relatively simple procedure, with a satisfyingly high degree of accuracy, was applied in this publication to impute the marital status of household members who were not DHS-eligible individuals.

The procedure starts with a data record for a target individual listed in the household roster whose marital status is unknown. It then uses information for this record regarding sex, age and relation to the household head, and searches for an individual who is a potential match among all those listed in the household roster. A potential match is an individual whose age and relation to head are such that the individual could be the target's spouse. Table A.I.1 shows the association between targets and other members of the household, defines the imputed values of marital status, and identifies the problems that emerge in each set of associations.

The simplest case occurs when the target individual is the head or the spouse of the head of household (rows 1 and 2 in table A.I.1). Whenever a target individual is the head of the household, the procedure searches for another individual listed among those in the roster whose relation to the household head is spouse. If one is found, both the target and the match individuals are classified as married. If none is found, the target individual is classified as unmarried. If the target is the spouse of the head, the situation is analogous and similar rules for imputing marital status are applied. Note that, according to these rules, all individuals living alone and who are not eligible for the main survey are classified as unmarried. This can obviously lead to incorrect decisions in situations when there is high prevalence of temporary absences of spouses due to migration, military service, population displacement, and so on. These types of errors, however, are very likely to be of less import for the older population, who are less likely to be in the labour force, less likely to be enlisted in the army and less likely to be temporary migrants.

If the target individual is not the head of the household, there are additional complications generated by one of the following two conditions: lack of correspondence and fostering arrangements. The third row of table A.I.1, for example, represents a situation where the target individual is the father (mother) of the head of the household. One would naturally search for an individual whose relation to head is mother (father). If one is found, then the imputation of marital status to the target is likely to be accurate. However, if one is not found, the imputation

"unmarried" may well be in error. This may occur if the target's mother (father) died or divorced and the father (mother) remarried some time later (but before the survey). It is only when the target's father's (mother's) current spouse is an adoptive or foster mother (father) and recognized as such by the variable "relation to head" that the procedure will actually estimate correctly the target's marital status. A similar case is represented in the fourth row of table A.I.1, where the target individual is the father-in-law or mother-in-law of the head of the household.

TABLE A.I.1. BASIC RULES FOR MATCHING TARGETS AND POTENTIAL SPOUSES

Target's relation to head	*Universe of potential matches*	*Imputation*	*Difficulties*
Head	Spouse head	Target: married Match: married	Temporary absences
Spouse of head	Head	Target: married Match: married	Temporary absences
Father/mother	Mother/father	Target: married Match: married	Temporary absences Remarriage
Father-/mother-in-law	Mother-/father-in-law	Target: married Match: married	Temporary absences Remarriage
Son/daughter	Son-/daughter-in-law	Target: married Match: married	Temporary absences Correspondences
Son-/daughter-in-law	Son/daughter	Target: married Match: married	Temporary absences Correspondences
Grandchild	Undefined	Not possible	No correspondence
Siblings	Undefined	Not possible	No correspondence
Other relative	Undefined	Not possible	No correspondence
Non-relative	Undefined	Not possible	No correspondence

Rows 5 and 6 in table A.I.1 illustrate cases where the sources of errors include lack of correspondence due to multiple possible associations between individuals. Given the usually coarse codes for the "relation to head" variable used in DHS (see the left-most column in table A.I.1), it is impossible to tell these associations apart and to identify an unambiguous marital status. Consider, for example, the fifth row of the table. In this case, the target individual is a son (daughter) of the head. If one finds a daughter (son)-in-law among individuals in the household roster, there will be no necessary correspondence between that person and the target, as the in-law listed in the roster may be the spouse of another sibling. This possibility can be ruled out, with a high degree of confidence, only in monogamous societies if the number of siblings in the roster equals the number of corresponding in-laws, that is to say, if the total number of siblings' spouses equals the number of siblings. However, if owing to temporary absences such a condition is not satisfied, at least one assignment to "unmarried" will be made in error.

Finally, there is the most problematic situation when the relation to head is coded "grandchild", "brother or sister", "other relative" or "not related". In this case, alternative solutions generate two different variants of the procedure. In the first variant, "unknown" marital status is assigned to all those individuals for whom relation to head is classified as belonging to one of the above four categories. In the second and less

conservative variant, the procedure uses information on age and sex to identify potential matches among those in the household roster whose relation to head is "other relative" or "not related". Individuals whose age and/or sex are unknown cannot be considered.

Accuracy of the imputation procedure

A direct, albeit partial, measure of the successful application of imputation, at least in the case of the first variant, is a low yield of undecided cases. In addition to persistent uncertainty, however, both variants of the imputation procedure may generate "false positives" (unmarried individuals whom the procedure imputes the status of married) and "false negatives" (married individuals to whom the procedure imputes status of unmarried). Table A.I.2 provides all the elements with which to define very simple measures of the accuracy of the procedure.

For variant 2, which does not allow for undecided cases (C = F = 0), the ratio S = A/M is analogous to a sensitivity ratio and measures the unconditional sensitivity of the imputation procedure, that is to say, the proportion of all individuals who are married whose marital status is correctly identified by the imputation procedure. In turn, the ratio Sp = A/(A+D) is analogous to a specificity ratio and measures the unconditional specificity of the imputation procedure, that is to say, the proportion of all individuals whose marital status is identified as married by the imputation procedure who are actually married. For variant 1, which does allow for unknown cases, the conditional sensitivity is given by the ratio A/(A+B) instead of the ratio A/M, while the conditional specificity is calculated in the same way as the unconditional specificity. The ratio of unknown cases to total cases, Z=(C+F)/(M+U), is a measure of the uncertainty remaining after imputation, whereas 1-Z is a measure of the total reduction of uncertainty attributable to the imputation procedure.

TABLE A.I.2. THEORETICAL DISTRIBUTION OF TARGET CASES

True marital status	*Imputed marital status*			
	Married	*Unmarried*	*Unknown*	*Total*
Married	A	B	C	M
Unmarried	D	E	F	U

To assess the accuracy of the imputation procedure, the imputation procedure was applied to the nine available DHS surveys that did elicit information on marital status of household members. The imputed and reported marital statuses were then compared. These surveys were heterogeneous in terms of region, culture and level of development. The imputation algorithm was applied as if marital status for individuals aged 60 years or over was unknown. The results of applying variant 1 of the procedure are summarized in table A.I.3, and those for variant 2 are summarized in table A.I.4.

TableA.I.3 shows that in all cases, specificity (Sp) attains values that are very close to 1. Sensitivity (S) is also generally high, ranging from 0.83 in Ghana to 0.98 in Turkey. These are satisfactorily high levels, especially if one considers that in each case, the values for S and Sp are both high. This observed pattern suggests that the higher sensitivity does not necessarily translate into a higher fraction of false negatives. Table A.I.3 also shows that the largest fraction of unknown cases (Z) is about 8 per cent in Yemen.

TABLE A.I.3. CROSS-CLASSIFICATION OF OBSERVED AND IMPUTED MARITAL STATUS FOR PEOPLE AGED 60 YEARS OR OVER AND MEASURES OF IMPUTATION ACCURACY: VARIANT 1

	Imputed marital status								
	Married		*Unmarried*		*Unknown*		*Total*		
True marital status	*Abs.*	*%*	*Abs.*	*%*	*Abs.*	*%*	*Abs.*	*%*	*Imputation accuracy*
Burkina Faso (1998)									S=0.97
Married	1 309	95.3	46	3.4	18	1.3	1 373	100.0	Sp=1.00
Unmarried	5	1.1	379	81.2	83	17.8	467	100.0	Z=0.05
Egypt (2000)									S=0.98
Married	3 257	96.0	74	2.2	62	1.8	3 393	100.0	Sp=1.00
Unmarried	13	0.6	1 948	94.2	106	5.1	2 067	100.0	Z=0.03
Ghana (1998)									S=0.83
Married	696	82.0	146	17.2	7	0.8	849	100.0	Sp=1.00
Unmarried	1	0.1	681	94.7	37	5.1	719	100.0	Z=0.03
India (1998)									S=0.96
Married	23 247	93.5	1 067	4.3	562	2.3	24 876	100.0	Sp=1.00
Unmarried	65	0.5	13 224	93.0	924	6.5	1 4213	100.0	Z=0.04
Pakistan (1990/91)									S=0.96
Married	2 317	93.7	94	3.8	63	2.5	2 474	100.0	Sp=1.00
Unmarried	9	1.0	847	91.8	67	7.3	923	100.0	Z=0.04
Turkey (1998)									S=0.98
Married	2 100	97.3	32	1.5	27	1.3	2 159	100.0	Sp=1.00
Unmarried	3	0.3	848	96.8	25	2.9	876	100.0	Z=0.02
Yemen (1992/93)									S=0.94
Married	3 803	90.7	256	6.1	135	3.2	4 194	100.0	Sp=1.00
Unmarried	2	0.1	1 285	79.7	326	20.2	1 613	100.0	Z=0.08
Colombia (2000)									S=0.90
Married	2 164	83.2	236	9.1	200	7.7	2 600	100.0	Sp=1.00
Unmarried	10	0.7	1 422	92.5	105	6.8	1 537	100.0	Z=0.07
Nicaragua (1997/98)									S=0.92
Married	1 939	88.2	159	7.2	101	4.6	2 199	100.0	Sp=0.99
Unmarried	16	1.0	1 363	89.1	150	9.8	1 529	100.0	Z=0.07

Source: Demographic and Health Surveys (DHS).
NOTE: Abs. = absolute number; S = sensitivity; Sp = specificity; Z = uncertainty.

TABLE A.I.4. CROSS-CLASSIFICATION OF OBSERVED AND IMPUTED MARITAL STATUS FOR PEOPLE AGED 60 YEARS OR OVER AND MEASURES OF IMPUTATION ACCURACY: VARIANT 2

	Imputed marital status						
	Married		*Unmarried*		*Total*		*Imputation accuracy*
True marital status	*Abs.*	*%*	*Abs.*	*%*	*Abs.*	*%*	
Burkina Faso (1998)							
Married	1 312	95.6	61	4.4	1 373	100.0	S=0.96
Unmarried	9	1.9	458	98.1	467	100.0	Sp=0.99
Egypt (2000)							
Married	3 274	96.5	119	3.5	3 393	100.0	S=0.96
Unmarried	16	0.8	2 051	99.2	2 067	100.0	Sp=1.00
Gana (1998)							
Married	696	82.0	153	18.0	849	100.0	S=0.82
Unmarried	2	0.3	717	99.7	719	100.0	Sp=1.00
India (1998)							
Married	23 419	94.1	1 457	5.9	24 876	100.0	S=0.94
Unmarried	90	0.6	14 123	99.4	14 213	100.0	Sp=1.00
Pakistan (1990/91)							
Married	2 340	94.6	134	5.4	2 474	100.0	S=0.95
Unmarried	10	1.1	913	98.9	923	100.0	Sp=1.00
Turkey (1998)							
Married	2 111	97.8	48	2.2	2 159	100.0	S=0.98
Unmarried	3	0.3	873	99.7	876	100.0	Sp=1.00
Yemen (1992/93)							
Married	3 838	91.5	356	8.5	4 194	100.0	S=0.92
Unmarried	4	0.3	1 609	99.8	1 613	100.0	Sp=1.00
Colombia (2000)							
Married	2 190	84.2	410	15.8	2 600	100.0	S=0.84
Unmarried	12	0.8	1 525	99.2	1 537	100.0	Sp=0.99
Nicaragua (1997/98)							
Married	1 948	88.6	251	11.4	2 199	100.0	S=0.89
Unmarried	20	1.3	1 509	98.7	1 529	100.0	Sp=0.99

Source: Demographic and Health Surveys (DHS).
NOTE: Abs. = absolute number; S = sensitivity; Sp = specificity.

The results displayed in table A.I.4 show that inclusion of unknown cases does not alter substantially any of the inferences regarding the sensitivity and specificity of the imputation algorithm. It suggests that, at least in cases where the fraction of undecidable cases is below 10 per cent, either of the two procedures can yield robust results.

Robustness of estimation using imputedmarital status

Although excellent performance of the aggregate measures of accuracy is a necessary condition for acceptability of the procedure, it is also important to verify that individual imputation leads to robust estimation of models involving other variables. Indeed, it is possible to obtain high sensitivity and specificity even when the errors in imputation of individuals' marital status translate, for example, into sizeable errors associated with estimated effects of other variables on marital status or of the latter on other variables. Whether this actually occurs will depend on the distribution of imputation errors across individuals or, alternatively, on whether or not they are truly random relative to other characteristics of interest.

In order to test this dimension of the imputation procedure, a conditional logistic model of the probability of living alone was estimated for individuals aged 60 years or over who were unmarried. The model was estimated using the imputed information about marital status and again using the reported marital status in those surveys that obtained the information on actual marital status. The independent variables included in the model were gender, age, education and place of residence (urban/rural). Table A.I.5 displays the odds ratios and corresponding standard errors associated with each of those variables. The first column displays the results from the model that used observed information on marital status. The second column displays results corresponding to the more conservative version of the imputation procedure (variant 1), that is, the one in which all undecidable cases are assigned the value "unknown". The third column displays the results obtained by using the less conservative imputation procedure, that is, the one in which a marital status is assigned to all individuals (variant 2).

There are two conclusions that we may derive from this table. First, in virtually all cases, the inferences about the direction and statistical significance of effects are identical, regardless of whether the observed or imputed marital status is employed. The second conclusion is that the differences produced by the two alternative imputation procedures are generally small; that is to say, the trade-off between the two procedures—increases in uncertainty relative to decreases in specificity and sensitivity—does not translate into an important impact in terms of robustness of inferences. The more conservative strategy for imputing marital status was applied in this publication.

TABLE A.I.5. ODDS RATIO FROM LOGISTIC REGRESSIONS OF LIKELIHOOD OF LIVING ALONE AMONG UNMARRIED OLDER PERSONS ON SELECTED VARIABLES ACCORDING TO THE FORM IN WHICH THE INFORMATION ON MARITAL STATUS WAS OBTAINED

	Information on marital status					
	Observed		*Imputed (version 1)*[a]		*Imputed (version 2)*[b]	
Variable	*Odds ratio*	*Std. error*	*Odds ratio*	*Std. error*	*Odds ratio*	*Std. error*
Colombia						
age1 (60-64)	0.781	0.179	0.913	0.174	0.853	0.158
age2 (70+)	0.890	0.172	0.761	0.128	0.842	0.138
female	0.449***	0.071	0.354***	0.049	0.435***	0.058
rural	1.644**	0.266	1.403*	0.201	1.289	0.180
some education	1.162	0.202	0.951	0.143	1.048	0.153
Egypt						
age1 (60-64)	0.622**	0.097	0.677**	0.100	0.664**	0.098
age2 (70+)	0.947	0.127	0.962	0.127	0.946	0.124
female	0.697*	0.100	0.604***	0.080	0.612***	0.080
rural	0.745*	0.086	0.732**	0.082	0.723**	0.081
some education	1.356*	0.194	1.418*	0.193	1.434**	0.193

TABLE A.I. 5 (*continued*)

Variable	*Information on marital status*					
	Observed		*Imputed (version 1)*[a]		*Imputed (version 2)*[b]	
	Odds ratio	*Std. error*	*Odds ratio*	*Std. error*	*Odds ratio*	*Std. error*
Ghana						
age1 (60-64)	0.910	0.231	0.913	0.211	0.948	0.217
ag 2 (70+)	1.697*	0.376	1.729**	0.356	1.812**	0.367
female	0.387***	0.076	0.333***	0.057	0.343***	0.058
rural	1.012	0.188	0.889	0.149	0.859	0.142
some education	1.275	0.293	1.240	0.248	1.281	0.253
India						
age1 (60-64)	1.020	0.088	1.020	0.083	1.022	0.083
age2 (70+)	0.831*	0.067	0.816**	0.062	0.816**	0.062
female	1.010	0.077	0.817**	0.055	0.863*	0.057
rural	1.304**	0.098	1.149*	0.080	1.163*	0.081
some education	1.005	0.079	1.035	0.074	1.050	0.075
Nicaragua						
age1 (60-64)	0.913	0.227	0.979	0.225	0.964	0.218
age2 (70+)	0.797	0.173	0.895	0.185	0.849	0.172
female	0.256***	0.044	0.250***	0.039	0.283***	0.044
rural	1.373	0.252	1.480*	0.251	1.518*	0.251
some education	1.724**	0.313	1.653**	0.280	1.661**	0.274
Pakistan						
age1 (60-64)	0.312*	0.156	0.280**	0.122	0.283**	0.123
age2 (70+)	0.501*	0.176	0.382**	0.123	0.385**	0.124
female	0.449*	0.145	0.344**	0.105	0.369**	0.113
rural	2.994**	1.092	2.339**	0.763	2.584**	0.840
some education	0.452	0.343	0.767	0.396	0.833	0.429
Turkey						
age1 (60-64)	0.623*	0.143	0.594*	0.133	0.619*	0.138
age2 (70+)	0.690*	0.127	0.688*	0.124	0.702*	0.126
female	0.707	0.142	0.713	0.138	0.704	0.135
rural	0.640**	0.105	0.654**	0.106	0.626**	0.101
some education	1.753**	0.301	1.865***	0.313	1.889***	0.315

Source: Demographic and Health Surveys (DHS).

NOTE*:* Significance level of odds ratios (whether ratio differs from 1.0): *p<0.05; **p<0.01; ***p<0.001.

[a] Version 1 of the imputation procedure excludes the "undecidable" cases, that is to say, those whose relation to head of household is "grandchild", "brother/sister", "other relative" or "not related".

[b] Version 2 of the imputation procedure attempts to seek matches for the "undecidable" cases based on the sex and age and relationship to the head of other members of the household; if not possible, individuals are imputed unmarried.