

Identifying Family Relationships for the World's Largest Collection of Census Microdata

IPUMS-International: Integrated Public Use Microdata Series

Matthew Sobek and Sheela Kennedy

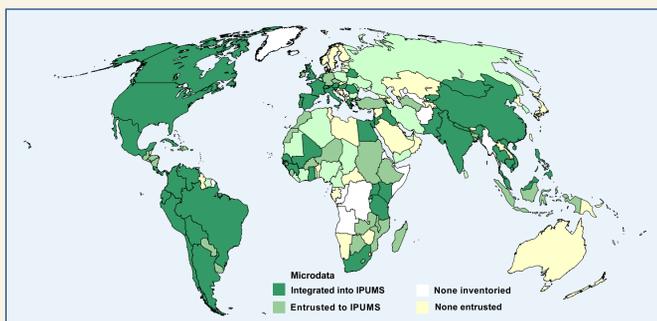
Minnesota Population Center, University of Minnesota

ABSTRACT

To facilitate comparative research on families and households, the IPUMS-International project has developed consistent "pointer" variables identifying each person's mother, father and spouse in a census microdata collection spanning 55 countries over five decades. These new variables provide researchers with a common tool for studying family interrelationships, removing the possibility that differing results are artifacts of different linking procedures.

This poster documents the methodology used to identify the most likely parent-child and spouse pairings. We show that the IPUMS pointers agree with direct reports of family interrelationships more than 98% of the time, and highlight for researchers factors that affect the precision of these links.

IPUMS-INTERNATIONAL PROJECT



What is IPUMS-International

- Freely available database of census microdata
- 325 million person records in 158 census samples from 55 countries, 1960 to the present
- Harmonized data and documentation

Countries in IPUMS-International

Africa	Asia	Americas	Europe
Egypt	Armenia	Argentina	Austria
Ghana	Cambodia	Bolivia	Belarus
Guinea	China	Brazil	France
Kenya	India	Canada†	Greece
Mali*	Iraq	Chile	Hungary
Rwanda	Israel	Colombia	Italy
Tanzania*	Jordan	Costa Rica	Netherlands†
South Africa	Kyrgyz Republic	Cuba*	Portugal
Uganda	Malaysia	Ecuador	Romania
	Mongolia	Mexico	Senegal*
	Nepal*	Panama	Slovenia
	Palestine	Peru*	Spain
	Pakistan*	Puerto Rico*	Switzerland*
	Philippines	Saint Lucia*	United Kingdom
	Thailand*	United States	
	Vietnam	Venezuela	

*Summer 2010 release

†Family interrelationship variables unavailable

FAMILY INTERRELATIONSHIPS

Family interrelationships in Census microdata

- Census data are typically organized hierarchically: individuals nested in households.
- Each person's relationship to a reference person in their household is usually known, but relationships to other persons are generally ambiguous.
- If these family relationships are identified, researchers can combine the characteristics of related and co-resident persons to create new variables and measures, for example:
 - » spouse's age
 - » mother's educational attainment
 - » own-child fertility

What are "pointers"?

- Pointers are locator variables—variables that identify each person's mother, father, or spouse, if one is present in the household
- The spouse locator variable identifies the person number of a person's spouse. (See example.)
- The mother and father locator variables "point" to a person's mother and father. (See example.)

Example: Family Interrelationships in a Census Household

Person Number	Relate	Age	Sex	Marital status	Children Ever-born	Spouse's Location
1	head	46	male	married	n/a	2
2	spouse	44	female	married	3	1
3	child	15	female	single	0	0
4	child	13	female	single	n/a	0
5	child	22	female	single	1	0
6	grandchild	3	male	single	n/a	0

Person Number	Relate	Age	Sex	Marital status	Children Ever-born	Mother's Location	Father's Location
1	head	46	male	married	n/a	0	0
2	spouse	44	female	married	3	0	0
3	child	15	female	single	0	2	1
4	child	13	female	single	n/a	2	1
5	child	25	female	single	1	2	1
6	grandchild	3	male	single	n/a	6	0

INFERENCE POINTERS

How are family relationships inferred?

- Family interrelationships are identified using relationship to head, marital status, fertility, and proximity in the household roster.
- IPUMS pointers are created simultaneously for all countries, applying the same method despite variation in data quality, availability, and household structure.

Empirical refinements

- Comparison with census pointers (direct reports of spouse or parent location)
- Using samples with detailed relationship data to test rules for less detailed samples

METHODOLOGY

General Design of Pointer Process

- Identify allowable relationship-pairings
- Set criteria for links within relationship pairing
 - » E.g. Age-differences, adjacency, marital status, fertility history
 - » Stronger criteria employed when relationship categories become more ambiguous
 - » Sample-specific rules when necessary

Search Process

Links between potential spouses are constructed first.

- We start by applying the strongest possible criteria for locating a spouse.
 - » The strongest criteria involve explicit relationship combinations such as head-to-spouse and parent-to-parent.
 - » Working sequentially downwards through a household, we evaluate each person in turn for a spouse link under these criteria.
 - » At the moment a person is linked they and their spouse are removed from further consideration.
- Subsequent passes use progressively weaker rules to make links—generally based on more ambiguous relationship pairings.

Child pointers are constructed next using a similar search process.

Rules for locating spouse or partner

1. Strong relationship pairing, couple adjacent
 - » E.g. head and spouse, or child and child-in-law
 - » Must be in a union
2. Strong relationship pairing, couple not adjacent
 - » Must be in a union
3. Weak relationship pairing, couple adjacent
 - » E.g. two "other relatives" or two non-relatives
 - » Marital status match preferred, tighter age requirements
4. Weak relationship pairing, couple not adjacent
 - » Marital status match preferred, tighter age requirements
5. Weak consensual unions pairings
 - » Unions between a non-relative and the head or any relative
 - » Must be in consensual unions, tighter age requirements
6. Sample-specific rules
 - » Unions between two children
 - » Exact marital status match, tighter age requirements

Rules for locating mother or father

1. Links involving head and spouse
 - » Minimal requirements, except when there is ambiguity in relationship codes (e.g. polygamy)
2. Links between children and grandchildren
 - » Tighter age requirements, preference known fertility
3. Links between other specified relatives
 - » Tighter age and marital status requirements
 - » Preference to persons with known fertility
4. Rules involving other unspecified relatives
 - » Tighter age and marital status requirements
 - » No links in excess of observed fertility
5. Rules between people unrelated to the head
 - » Tighter age and marital status requirements
 - » No links in excess of observed fertility
 - » Must be adjacent

EVALUATION

IPUMS pointer accuracy

We use census pointers (direct reports of spouse and parent line numbers) to evaluate the final IPUMS pointers for the overall rate of disagreement (see table) and factors affecting accuracy (see conclusions).

Difference between IPUMS and Census Pointers (%)

Census	Spouse	Mother
Armenia 2001	1.3	1.1
Belarus 1999	0.2	0.3
Brazil 1991		0.5
Portugal 1981	0.3	1.1
Portugal 1991	0.2	1.9
Portugal 2001	0.2	0.6
Romania 1977	0.4	0.4
Romania 1992	0.5	0.4
Romania 2002	0.1	0.2
South Africa 2001	1.2	4.9
South Africa 2007	0.8	3.9
Spain 1991	0.1	
Spain 2001	0.2	0.3
TOTAL	0.5	1.3

CONCLUSION

- Across samples with census pointers, the IPUMS pointers are in close agreement:
- 99.5% for the spouse pointer and 98.7% for parent pointers, although disagreement rates are higher for individual countries.
 - Characteristics of spouse and parent-child pairings produced by the IPUMS pointers resemble closely those of census pairings.

The agreement rate falls when there are multiple spouse or parent candidates or when the data available to evaluate potential pairings is limited.

For example:

- Complexity of household structure; polygamy
- Availability of fertility data
- Availability of detailed relationship categories (e.g. grandchildren or in-laws)
- Meaningfulness of household order

ACKNOWLEDGEMENTS

Support for this research was provided by the National Science Foundation and the National Institute of Child Health and Human Development. We are grateful to Vladimir Vladykin and Monty Hindman for invaluable technical and design assistance.