

Identifying Family Interrelationships in the World's Largest Census Microdata Collection

Matthew Sobek and Sheela Kennedy

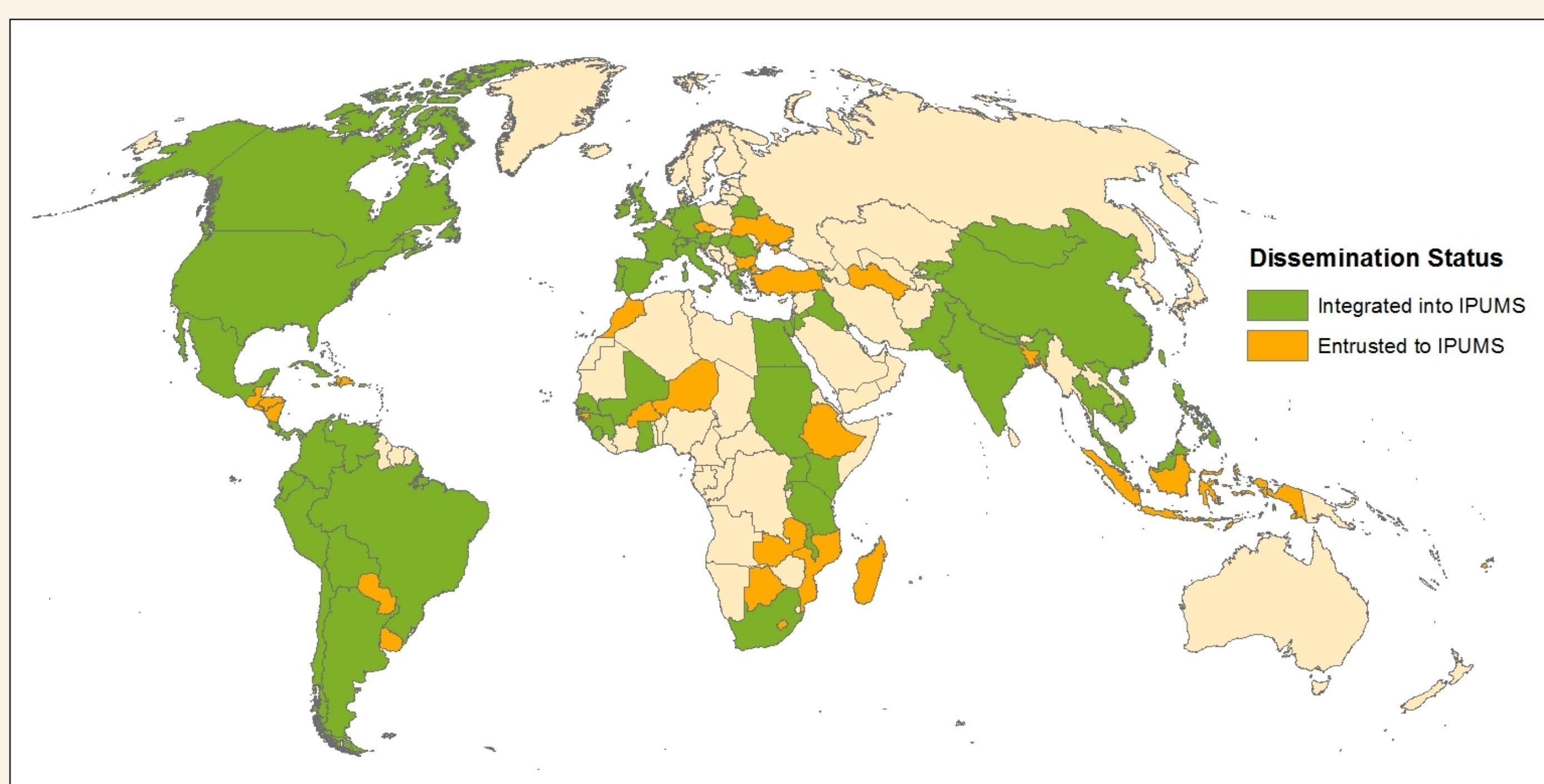
Minnesota Population Center, University of Minnesota

SUMMARY

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 62 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.

IPUMS-INTERNATIONAL PROJECT

IPUMS-International is a freely available database of census microdata from the 1960s to the present. It currently contains 397 million person records from 185 censuses in 62 countries. Data and documentation are harmonized across samples.



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 identifies the person number of a person's spouse.
- The mother and father locators "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	22	female	single	1	2	1
6	grandchild	3	male	single	n/a	5	0

ACKNOWLEDGEMENTS

Support for this research was provided by the U.S. National Science Foundation and the National Institute of Child Health and Human Development.

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
 - Use stronger criteria when relationship categories are more ambiguous
 - Apply sample-specific rules when necessary

Search Process

- Links between potential spouses are constructed first.
- Start by applying the strongest possible criteria for locating a spouse: explicit relationship combinations such as head-to-spouse and parent-to-parent.
 - Working sequentially downwards through a household, evaluate each person in turn for a spouse link under these criteria.
 - When 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

- Strong relationship pairing, adjacent (e.g., child and child-in-law)
- Strong relationship pairing, not adjacent
- Weak relationship pairing, adjacent; require tighter age requirements
- Weak relationship pairing, not adjacent
- Weak consensual union pairing (e.g., non-relative to a relative of head)
- Sample-specific rules (e.g., unions between two "children")

Rules for locating mother or father

- Links involving head and spouse: minimal requirements except for polygamy
- Links between children and grandchildren
 - Tighter age requirements, privilege links to persons with known fertility
- Links between other specified relatives
 - Privilege links to persons with known fertility
- Links involving other unspecified relatives
 - No links in excess of observed fertility
- Links between people unrelated to the head
 - 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 overall rate of disagreement (see table) and for factors affecting accuracy (see conclusions).

Difference Rate between IPUMS and Census Pointers (%)

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