# *Multivariate Analysis of the TFR and Its Components Using Census or Household Survey Data: Methodological Improvements*

**R. D. Retherford, H. Eini-Zinab,**

**N. Ogawa, and R. Matsukura**

# *What is meant by "TFR and its components"?*

- Old methodology (last census conference):
  - Parity progression ratios (PPRs), mean and median ages at first marriage, mean and median closed birth intervals, $TFR_{ppr}$, and $TMFR_{ppr}$
- Improved methodology:
  - All the above components, plus ASFRs, $TFR_{asfr}$, and mean and median ages at childbearing (both overall and by child's birth order)
  - **Note:** In the improved methodology $TFR_{asfr} = TFR_{ppr}$

# *Both old and improved multivariate methodology:*

- Methodology applies to both synthetic cohorts (period data) and real cohorts (cohort data)
  - E.g., period TFR and cohort TFR
- Underlying methodology is complex, but final output is simple bivariate tables
  - E.g., tabulation of TFR by categories of education with urban/rural residence controlled (i.e., held constant --- shown in next slide

## *Unadjusted and adjusted values of the period TFR by education, Philippines 1993-97*

| | | |
|---|---|---|
| Low | Unadjusted | 4.67 |
| | Adjusted | 4.30 |
| Medium | Unadjusted | 3.73 |
| | Adjusted | 3.69 |
| High | Unadjusted | 2.81 |
| | Adjusted | 2.88 |

"Unadjusted" means "without controls". "Adjusted" means "with controls". Urban-rural residence is controlled in the adjusted estimates. Results based on Philippines 1998 DHS. Old methodology used in this table. (Had new methodology been used, the table would still have the same simple format, but the numbers would be a little different.)

# *Required data for both old and new methodology*

- **The census or household survey must include a question on number of children ever born (CEB)**
  - CEB Preferably asked of women out to the age of 65, so one can compute period estimates of TFR and its components for each of the last 15 years (15 years ago women age 65 were age 50)
- Desirable but not essential:
  - A question on age at first marriage (in order to get estimates of nuptiality and marital fertility)

## *Also essential:*

- A complete listing of household members by age, sex, marital status, and relation to household head (or householder)

# *First step: Reconstruction of complete birth histories*

- **A question on number of children ever born (CEB) is required**
- **Start with incomplete own-children birth histories**
  - "Own" means "matched to a mother in the household"
  - Child's age translates into year of birth
  - Dead children and children living elsewhere are missing from a woman's own-children birth history
  - Number of missing children is calculated as CEB minus number of own (i.e., matched) children
  - Missing children are imputed into the birth history using probabilistic procedures developed almost two decades ago by Norman Luther
  - **Result:** Reconstructed birth histories

# *How accurate are the reconstructed birth histories?*

- In general, they are not accurate for any particular woman
- But when estimating aggregate measures, errors in individual birth histories mostly cancel out
  - Estimates of TFR and its components are reasonably accurate
  - **Caveat:** Age reporting and information on CEB must be reasonably accurate.

# *TFR and its components are calculated from the birth histories*

- Griffith Feeney developed PPR-based methods (not multivariate) for doing this
  - Focused on period estimates of PPRs and $TFR_{ppr}$ for each of the past 15 years
  - **Note:** The birth histories can be either real (as in DHS surveys) or reconstructed (as in censuses and household surveys
- The new methodology presented here makes Feeney's method multivariate
  - E.g., TFR by categories of education, controlling for other socioeconomic predictor variables

# *Modeling approach*

- Discrete-time survival models
  - Either discrete-time logit or complementary log-log
  - They give close to identical results
  - **Note:** A discrete-time survival model may be viewed as a multivariate life table
- One model for each parity transition
  - B-M (woman's own birth to first marriage)
  - M-1 (first marriage to first birth)
  - 1-2 (first birth to second birth)
  - And so on (with an open parity interval at the end)

# *General features of the model for a particular parity transition*

- Response variable in model for a particular parity transition (*i* to *i*+1):
  - *P* (either a first-marriage probability or a "next-birth" probability --- like an age-specific probability of dying $q_a$ in an ordinary life table)
  - First marriages treated mathematically like next births
- Predictor variables:
  - *a* (woman's age) and *t* (duration in parity), plus socioeconomic predictor variables of interest (e.g., education and urban/rural residence)

## *Transition probabilities $P_{ait}$ yield a "global life table" covering all parity transitions*

- $P_{ait}$ denotes the set of transition probabilities by age, parity, and duration in parity for each possible combination of values of the socioeconomic predictor variables
  - Global life table is calculated from the $P_{ait}$
  - Global life table has three basic dimensions (*a*, *i*, and *t*)
  - Starting at age 10 or 15, women are followed one year at a time by age, parity, and duration in parity until they reach age 50

## *Events (first marriages or next births) in the global life table are denoted by $f_{ait}$*

- The global life table generates values of $f_{ait}$ (like age-specific deaths $d_a$ in an ordinary life table)
- Values of TFR and its components are calculated using the global life table values of $f_{ait}$

# *The global life table is multivariate*

- It follows that all measures calculated from the global life table (TFR and its components) are also multivariate
  - E.g., one can get estimates of TFR (or one of its components) by education, with other socioeconomic variables controlled

# *The global life table is internally consistent*

- All measures calculated from the global life table are consistent with each other
  - That is why $TFR_{asfr} = TFR_{ppr}$ when calculated from the global life table

# *Why do this?*

- Get closer to true effects of predictor variables
  - Potentially confounding variables are controlled
- As fertility comes down, there will be less money for fertility surveys
  - Increasing need to utilize censuses/household surveys that collect data for other purposes
  - New methodology yields a great deal of information about marriage and fertility levels, trends, and determinants, without huge cost of new data collection

# *Why do this? (continued)*

- With large census samples, one can also do separate analyses for local areas (e.g., provinces)
  - This is policy-relevant
  - Population policies often administered at the province level

# *Household surveys*

- E.g., labor force surveys, income and expenditure surveys, American Community Survey
- Need to add a few extra questions
  - Complete household listing by age, sex, marital status, and relation to head (essential)
  - CEB (essential)
  - Age at first marriage (highly desirable)
- Don’t need these extra questions every survey round
  - Once every 5 or 10 years would be enough

# *How well does the methodology work?*

- Old multivariate methodology with *i* and *t* (but not *a*) already works well
  - Spelled out in detail in an EWC working paper
  - Short version in a forthcoming paper in *Demography*
  - Papers based on Philippines DHS surveys using real real birth histories

# *Improved methodology (with a, i, t) also works well*

- Tests based on real birth histories from Philippines DHS surveys
  - Close agreement of model-predicted estimates of TFR and components with standard birth history estimates of the same measures
- Application to reconstructed birth histories yet to be done
  - Involves modification and packaging of complex programs for birth history reconstruction
- No publications yet, using improved methodology

## *Computer programs will be placed in the public domain*

- They will be well-documented and in some cases packaged with an easy-to-use graphic interface
- They will be downloadable from the East-West Center website
- We hope that national census and statistical offices will use the programs!

## *References (old methodology that incorporates parity and duration in parity but not age):*


- Retherford, R. D., N. Ogawa, R. Matsukura. and H. Eini-Zinab. 2008 (revised October 2008). *Multivariate Analysis of Parity Progression-Based Measures of the Total Fertility Rate and its Components Using Individual-Level Data*. East-West Center Working Papers, Population and Health Series, No. 119. Honolulu: East-West Center. (Downloadable at http://www.eastwestcenter.org/stored/pdfs/POPwp119.pdf)
- Retherford, R. D., N. Ogawa, R. Matsukura, and H. Eini-Zinab. Multivariate analysis of parity progression-based measures of the total fertility rate and its components. Forthcoming in *Demography*. (Short version of Working Paper 119 above).