Using census data to create a detailed database for fertility analysis

Rebecca Kippen and **Peter McDonald**

Australian Demographic and Social Research Institute
The Australian National University

The standard period measure of fertility, the Total Fertility Rate, takes account (standardizes) only for the ages of women. In reality, age is a relatively weak predictor of whether or not a woman will give birth in a given year and, accordingly, projections of births that rely only on age-specific birth rates have performed badly across history. We have argued that more precise measurement of period fertility and projection of future births should involve simultaneous controls for the age of the woman, her parity and the length of the interval since her last birth. We have found, at least for Australia, that the probability of giving birth for a woman aged x with y existing children (where y>0) who had her previous birth z years ago is very stable over a long period of time. The advantage that this relative stability implies for the projection of births should be utilized (McDonald and Kippen 2007, 2008a, 2008b).

It is rare for all three of these characteristics to be controlled simultaneously because the data have been unavailable or the effort involved in obtaining the data has been prohibitive. Furthermore, we have shown that it is important to work with data by single years of age, single parity and single year of duration since the previous birth. This implies a matrix of some 1500 cells of values for each year. Accordingly, conventional birth histories available from sample surveys are not large enough to provide reliable data. A complete birth registration system in which these three characteristics are available could provide the necessary information on births (the numerator) but it does not provide the information ton the population that is required (the denominator). In the Australian case, the birth registration system does not provide information on these three characteristics. In relation to the population at risk, while parity data are collected at some Australian censuses, the interval since the most recent birth is not collected. Furthermore, this information is not collected in complete form in the Australian vital statistics collections. Thus, we have developed methods to obtain the required data from complete counts from censuses, specifically from the 1981, 1986, 1991, 1996, 2001 and 2006 Australian Censuses of Population and Housing. This paper describes the procedures that we use.

For each census, as with the Own-Children Method, Australian-resident women in each household are matched to their own children living in the same household at the time of the census. This is done using the ‘relationship in the household’ census variable, which describes the relationship of each person in the household to the household reference person (Person 1 on the census household form or Person 1 and Person 2 if it is a couple family). The Australian censuses identify whether a person in the household is the natural child of Person 1 or Person 2 and, where it is a couple household, the couple are required to be listed as Person 1 and Person 2. Where a child’s parent is not Person 1 or Person 2, a relatively rare circumstance, a set of rules was determined to link the child to the person most likely to be his or her mother.

Available characteristics from these data are age of woman, number of own children in the household, and age of each of her own children resident in the household at the time of each census. If all the women's children are in the household and the link between mother and own children is perfect, with some manipulation, this is equivalent to a woman's birth history. We first make the assumption that the data are a perfect representation of the woman's birth history and then correct for errors later. So the census 'birth history' data are used to construct two sets of tables for each census. The first set consists of counts of women by age by the age of the oldest child by the age of the second child by the age of her third child, and so on – a cross classification of the age of the mother by the ages of each of her children simultaneously. The second set consists of counts of women by age with j or more own children, classified by the age of child j by the age of child j–1.

The first set of data is used to construct distributions of women by single year of age by number of own children in the household for each year 1980–2006. For example, the distribution for 2001 is calculated by tabulating women by number of own children from the 2001 census count. The distribution for 2000 is calculated by taking the 2001 distribution, subtracting one year from the age of each woman, and, in addition, adjusting down by one the number of her own children if she had a child aged zero years at the 2001 census. The distribution for 1999 is calculated by 'younging' women in the 2000 distribution, and subtracting one from the number of own children for every child aged one year at the 2001 Census, and so on back to 1996, the year of the previous census. The same process is then repeated for the other censuses. No attempt is made to apply mortality rates in this 'reverse survival'. Thus, no attempt is made to place dead children into the woman's birth history. Our view is that this is better accomplished in the Australian case where child mortality is low by using the adjustment techniques that we describe below.

This 'reverse survival' results in five sets of annual distributions: 1981–86, 1986–91, 1991–96, 1996–2001 and 2001-2006. The use of censuses five years apart means that, for the most part, we are dealing only with young children, those under five years of age. In Australia, it is very rare that a child under the age of five years is not living with its natural mother. This almost negates the issue of non-own children, that is, children who are not able to be linked to their natural mothers. Again, we consider that the adjustment techniques described below are adequate and no specific adjustment for non-own children is required.

These reverse survived populations are used to calculate, for each year 1981–2000, the probability that a woman aged x in year y with j own children will progress to j+1 children in ageing to x+1 years in year y+1.

The number of own children in the household is not equal to the woman's parity because children may have died or left the maternal home. The transition probabilities can be adjusted so that they refer to parity rather than to own children. This is done by comparing the own-children transition probabilities for each year 1991–2000 to the parity transition probabilities for 1991–2000 calculated previously by Kippen (2003). We find that the ratio of transition probabilities for each age of woman and own children/parity is close to 1 and relatively constant across the ten years of data. We therefore use these ratios averaged across the decade to inflate the own-children transition probabilities to obtain cohort parity progression probabilities for the progressions to first, second, third, and fourth birth.

These parity progression probabilities—the probability that a woman aged x in year y with parity j shifts to parity j+1 in ageing to x+1 years in y+1—are used to calculate parity distributions for each age of woman 15–49 and each year 1981–2006, beginning with the parity distribution calculated from the 1981 census (which asked 'For each woman, how many babies has she ever had? Do not include still-births'). As a check for accuracy, the calculated age-parity distributions for 1986, 1996 and 2006 are compared to those derived from the 1986, 1996 and 2006 Censuses (which also asked women their parity). We find in all cases that the matches are very close. The implied annual age-specific fertility rates are also compared to those derived from birth registers. Again they match very closely.

The annual parity distributions and transition probabilities are used to calculate annual births by birth order and age of mother assuming cohorts of 100,000 women at each age, 1981–2006. Births of order 2, 3, 4 for each year are then classified by duration since previous birth. This is done using the second data set of data from each census of counts of women by age with j or more own children, classified by age of child j by age of child j–1. For example, there are 3,658 second births to women in ageing from 30 years in 1982 to 31 years in 1983. The distribution of these second births by interval from first birth can be calculated by considering women aged 34 years in the 1986 census with two or more children, second child aged 3 years, by age of first child. The age interval between the second and first child is the relevant birth interval.[1]

Synthetic transition probabilities are then calculated:

$$q_{j,x,y,i} = \frac{P_{j,y-i,x-i;j+1,x,y}}{P_{j,x-i,y-i} - \sum_{n=1}^{i} P_{j,x-i,y-i;j+1,x-n,y-n}} ,$$

1 We find that there is no need to adjust birth intervals to take account of children missing from the household. A comparison of interval distributions for all women, and women for whom number of children is equal to parity, for the 1981, 1986 and 1996 censuses shows that they are virtually identical. This suggests that if mothers have children missing from their household, it is likely that all their children are missing, or that the oldest or youngest are missing.

$1 \le j \le 3$,
$15 \le x \le 48$,
$1981 \le y \le 2000$,
$0 \le i \le 14$,

where $q_{j,x,y,i}$ is the probability that a woman aged x in year y of parity j with her jth birth i years ago progresses to birth j+1 in ageing from x years in year y to x+1 years in year y+1,
$P_{j,x,y}$ is the number of women who had a birth of order j between age x and x+1 and year y and y+1, and
$P_{j,x-i,y-i;j+1,x,y}$ is the number of women who had a birth of order j between age x–i and x–i+1 and year y–i and y–i+1 and a birth of order j+1 between age x and x+1 and year y and y+1.

The data collection provides us with full information on fertility achievement: the distribution of the female population of Australia by age, parity and interval since last birth and annual births with the same description for every year from 1980 to 2006. It then becomes a straightforward exercise to obtain birth probabilities simultaneously by single years of age, single parities and single years of duration since the last birth. These probabilities and the most recently available population at risk can then be used to make projections of future births. The improved accuracy of these projections is not only the result of the more detailed probabilities that are used. Importantly, it is also the result of taking into account the more detailed structure of the population at risk. This structure contains a 'momentum' deriving from past history that is vital (and very accurate) in projecting births of order two and over. It does not eliminate the distortions created by 'tempo effects' because it does not provide any better method of projecting first births, the main driving force of tempo effects. We consider that trends in the probability of first births must be modelled using social behavioural factors, but, once this is done, the trends in second and higher order births are very accurately projected using only the detailed demographic data base. We have applied this method to project births in Australia from 2001 to 2006 with a comparison against births that actually occurred in this period. The projection accurately projected a turning point in the Australian Total fertility Rate during these years (McDonald and Kippen 2008b). We have also used the data base to examine how changes in the timing of births affect birth rates (McDonald and Kippen 2008a).

Kippen, R. 2003. 'Trends in age- and parity-specific fertility in Australia,' *Working Papers in Demography*, No. 91, Demography and Sociology Program, The Australian National University, Canberra.

McDonald and Kippen, R. 2007. 'The Intrinsic Total Fertility Rate: A new approach to the measurement of fertility'. Paper delivered to the Population Association of America Meeting, New York.

McDonald P. and Kippen, R. 2008a. 'Delay of first births in Australia, 1981-2001: Its nature and its impact upon the timing and incidence of second births'. Paper delivered to the Annual Meeting of the Population Association of America, New Orleans.

McDonald P. and Kippen, R. 2008b. 'The projection of births: a new approach'. Paper delivered to the Conference of the Australian Population Association. Alice Springs, July.