CHAPTER

16

Natality

Measures Based on Censuses and Surveys

THOMAS W. PULLUM

The previous chapter discussed the use of vital statistics to measure natality, especially for areas with established and relatively complete vital registration systems. This chapter continues the discussion of natality, but by use of census and sample survey data. Methods and issues particularly relevant to the measurement of natality in statistically less developed areas will be the subject of Chapter 22. Reproductivity, in which survivorship is combined with fertility, is the topic of Chapter 17. Some measures are discussed in more than one of these chapters because they can be estimated from various sources or are conceptually closely related. This chapter focuses on how to obtain fertility measures from a census or survey data, even if the measures can also be calculated from vital statistics data. When useful, comparisons will be made with a wider range of measures.

Some examples will illustrate the nature of the overlap among these chapters. The classical estimate of the crude birthrate (CBR) uses vital statistics data for the numerator and census data for the denominator, combining generic types of data from both Chapters 15 and 4. As we will see, it is also possible to estimate the CBR entirely from a fertility survey if it includes a household roster. Thus, there is some discussion of the CBR here, as well as in Chapter 15. The total fertility rate (TFR) can be calculated completely with survey data and is similar to the gross reproduction rate (GRR) and the net reproduction rate (NRR), measures of reproductivity discussed in Chapter 17.

Most of the measures in Chapters 15, 16, and 17 can be classified into three types, according to their numerators and denominators. *Birthrates* have both male and female births in the numerators and both males and females in the denominator. The crude birth rate (CBR) is an example of a birthrate; indeed, it is the only measure of natality consistently labeled a birth rate. By contrast, *fertility* rates have both male and female births in the numerators, but just one sex in the denominator, usually females because fertility has traditionally been considered to be an attribute of women, and most data sources for births provide more information about the mother than about the father. The general fertility rate (GFR) and TFR are the most common examples of fertility rates. Third, a *reproduction* rate is limited to female births in the numerator as well as females in the denominator, and describes the replacement of females by females (or males by males). The GRR and NRR are the best known examples of reproduction rates.[1]

The distinction between birth, fertility, and reproduction rates is didactically useful and the computational relations between them can be seen as simple. For example, the GFR can be calculated by dividing the CBR by the proportion of the total population who are women aged 15 to 44. The GRR can be calculated by multiplying the TFR by the proportion of all births that are girls (assuming that the proportion of births that are girls does not greatly depend on the age of the mother).

Recent decades have seen a major transformation in the kinds of data available for demographic estimation, particularly related to natality. When this book originally appeared, it made only limited reference to fertility surveys and the birth rosters they contain. Most examples in the original chapter were based on census data. Since that time, there has been a complete reversal in the relative importance of censuses and surveys for measuring fertility. In the United States, the National Center for Health Statistics has conducted several rounds of the National Survey of Family Growth (often referred to as "NSFG"). Internationally, with primary sponsorship by the U.S. Agency for International Development, nearly two hundred fertility surveys have been conducted by the World Fertility Survey (WFS, 1973–1984) and the Demographic and Health Surveys

[1] This classification differs from common usage and the terminology of the IUSSP's *Multilingual Demographic Dictionary*, which uses "birth" and "fertility" interchangeably for some natality measures, particularity age-specific and order-specific rates and probabilities.

(DHS, since 1985). Many other surveys have been carried out with support from the U.S. Centers for Disease Control and Prevention (CDC) or under other auspices.

Along with this revolution in data availability, recent decades have seen a shift from aggregates to individuals as the units of analysis, and the corresponding adaptation of statistical tools to the analysis of demographic data. Statistical methods themselves have expanded enormously in recent decades because of a contemporaneous growth of computing capacity. Thirty years ago, for example, Poisson regression, logit regression, and hazard models, all of which are particularly appropriate for demographic analysis, were unavailable. For these reasons, this revised chapter differs dramatically from the original version.

We will present the measurement of fertility as fundamentally a description of an individual-level process. When at all possible, fertility measures are now generated with individual-level data—a fertility survey or a public-use sample from a census—rather than the tabulated information in a full census. The most useful measures of fertility, and indeed most of the traditional measures, can be interpreted in one or the other of the following two ways: (1) as the average or expected number (in the statistical sense) of births that a woman has in an interval of time, "controlling" for such characteristics as her age, marital duration, and parity, or (2) as the probability that a woman has a birth in an interval of time, controlling for such characteristics as her age, marital duration, and parity.

Otherwise, these measures differ from one another only in terms of the reference interval of time, what they "control" for, how they control for it, and whether they are cohort or synthetic cohort measures. When calculated from a sample, each measure has a standard error (sometimes difficult to estimate) that can be used to construct confidence intervals or test hypotheses. Most have the potential to be included in some form of multivariate analysis.

TYPES OF DATA AVAILABLE IN A CENSUS OR SURVEY

Census Data

Censuses have long been effectively employed in the more developed countries in the measurement and analysis of fertility, especially in the "dimensions of time and space." Census data on children ever born and the age-sex distribution have been used to track and analyze historical changes in fertility, including the years before vital registration was initiated or adequately developed for demographic applications. For example, the U.S. census data on children ever born, as reported for elderly women, have been employed to analyze the historical shifts in the familial support available to them. In addition, censuses have been used to describe and analyze geographic variations in fertility within countries, both currently and historically. Censuses providing retrospective data on fertility, released as public use microdata samples, permit manipulation of individual-level data that can be linked to other demographic and socioeconomic variables for current and historical fertility analysis. For a further discussion of measures of fertility based on aggregated census data, refer to Chapter 22 of this volume (statistically less developed areas) and to Chapter 17 of the first edition of this book (H. S. Shryock, J. S. Siegel, and E. G. Stockwell, 1976) (more developed areas).

As described elsewhere (see, for example, Chapter 2), many countries of the world conduct a census every 10 years, and some do so every 5 years. Prior to the 1970s, a census typically provided the best available data for estimating fertility, particularly in a less developed country. Vital registration systems were (and often continue to be) seriously incomplete, and only a few countries had conducted large-scale fertility surveys. This data deficiency led to creative ways to estimate fertility indirectly using one census or, even better, two successive censuses.

Some censuses include information on the number of children ever born (CEB). For example, this was a standard item for U.S. censuses from 1900 through 1990, but it is no longer collected in the U.S. census. This item is useful but says nothing about the timing of the births, apart from inferences based on the woman's age or the ages of children in the household.

Prior to the widespread use of fertility surveys, the inadequacy of vital registration systems in less developed countries led to the inclusion of fertility-related questions on census forms. This occurred notably in Africa, where William Brass of the London School of Hygiene and Tropical Medicine advised several central statistical offices, but also in Latin America and Asia. Such questions may ask whether a child was born in the last year (or another reference period), the number of children born in the previous 5 years (or another reference period), or the length of time since the most recent birth. The value of such data depends on a correct interpretation of reference periods and time intervals.

As a minimum, every census produces an age distribution. This contains information about fertility because the people observed to be age a at last birthday are the survivors of persons born exactly a to $a + 1$ years before the census. For example, infants, who are under one year of age (aged zero) at last birthday, are the survivors of the births during the previous elapsed year (not calendar year). The number of surviving children is less than the number of births (assuming no net immigration), so it is necessary in estimating the number of births with such data to use a reverse survival method, requiring assumptions about mortality. Many censuses in East and Southeast Asia include information to identify the children who were born to a specific woman in the household. With a plausible life table that spans the ages of the children, and also the mothers, it is possible to use the "own-children" method to estimate age-specific fertility rates during the 15 or so years before the census. This method falls within the rubric of indirect estimation, which is discussed elsewhere, including Chapter 22 and Appendix C.

The original version of this chapter devoted substantial space to the child-woman ratio (CWR), that is, the ratio of the number of children aged 0 to 4 to the number of women aged 15 to 49 reported in a census. The CWR is calculated from an age-sex distribution. If mortality is ignored, it is approximately five times the general fertility rate (discussion follows). In view of the current availability of more appropriate data, we have less need for the CWR as a measure of fertility and will discuss it only briefly.

In summary, census data alone are not as well suited for fertility measurement and analysis as survey data and now are infrequently used for this purpose, apart from the own-children method. Censuses provide population counts, or are the basis for intercensal and postcensal estimates of population, that can serve as the denominators of some rates; but they generally provide useful information about numbers of births only if supplemented with estimates of survivorship.

Survey Data

Special surveys on fertility, contraceptive use, and related topics are now the primary source of data for fertility analysis. Prior to the early 1970s, when the World Fertility Survey (WFS) began, most surveys of this genre were oriented primarily around the estimation of contraceptive prevalence. When WFS was launched, there was considerable skepticism that reliable retrospective birth histories could be collected in the less developed countries. Many experts believed that the reported birth histories would be incomplete because respondents would omit children who had died or were born long ago, and birth dates would often be unknown or erroneously reported. The WFS surveys soon demonstrated that reliable birth histories could indeed be collected.

In a typical fertility survey, a female respondent is first asked how many children she ever had, and how many are still alive, and then she is asked a series of questions (mainly date, sex, and survivorship) about each birth, beginning with the most recent one and working backward. (Some surveys begin with the earliest birth and work toward the present.) In addition, fertility surveys include the woman's own date of birth and a marriage history, giving dates of marriage and of marital dissolution. The definition of marriage is generally flexible and includes cohabitation. Many of the early surveys defined eligible respondents to be ever-married women aged 15 to 49, but it is now more common to include all women 15 to 49 (occasionally 15 to 44), regardless of marital status. Most surveys include a household roster that lists all persons in the household (with some characteristics, such as age, sex, and relation to head), including households that have no eligible respondents. This roster is especially important for calculating all-women fertility rates if the eligible respondents are limited to ever-married women.

The information in the birth history is coded onto a computer record, or set of records, for each case. To facilitate calculations, dates consisting of a month field and year field are typically converted into century month codes (cmc's), which begin with cmc = 1 for January 1900. Thus, for example, the cmc for August 1999 would be $8 + 99*12 = 1196$, because August is the eighth month of the year and a year contains 12 months. It seems likely that cmc's will continue to originate with January 1900 well into the 21st century. The date of interview is also converted to a century month code. This date will vary over the interval of data collection, usually several months. Care must be taken with the calculation of rates that extend into these months, because respondents provide incomplete information about the interval of data collection. Days of the month are ignored; in calculations it is necessary to make some arbitrary but consistent assumptions, for example, that events always occur on the first day of a month.

A birth history, sometimes requiring reference to the woman's birth date, marriage history, and date of interview, thus provides the following kinds of information for each respondent:

- Children ever born and children still living (forced to match the responses to the direct questions on these totals)
- Number of births in a window (interval) of age, calendar time, time since survey, or marital duration (elapsed time since first marriage), or an intersection of such windows
- A classification of these numbers of births into whether they were marital, premaritally conceived, or premaritally born, using either the "ever-married" or "currently married" criterion for marital status
- Length of intervals between births, including the interval from first marriage to first birth (closed intervals) and from first marriage (if no birth has occurred) or latest birth to date of interview (the "open" interval)
- Exposure to risk of an event, as described in the next section

The respondents can also be classified according to other covariates for which data may be available, such as education. With this information it is possible to calculate rates within subgroups or to use a multivariate method.

This chapter includes examples from a specific Demographic and Health Survey, the 1998 National Demographic and Health Survey of the Philippines, the main results of which were published in January 1999 (Philippines National Statistics Office and Department of Health, and Macro International, 1999). It is subsequently referred to as "Philippines

NDHS Report, 1999". The fieldwork for this survey was conducted from early March through early May 1998 and included interviews with 13,983 women. The rates given here differ somewhat from those in that report, mainly because we simplify the illustrative calculations by not using weights. A substantive analysis of this survey should use weights, as in the published report.

MEASURES OF FERTILITY IN AN INTERVAL OF AGE OR MARITAL DURATION

This section discusses the most common of the specific fertility rates and birth probabilities. All calculations are based on individual-level data from a survey.

Terminology

In fertility analysis, as in the analysis of mortality, migration, or other demographic events, *exposure to risk* is a key concept. This term refers to the time during which a person is at risk of experiencing an event, whether or not the event actually occurred. For example, the age-period-specific fertility rate for ages 20 to 24 in 1990–1994 is estimated from women who were in at least part of the age interval 20 to 24 during at least part of the time interval 1990–1994. A particular woman's exposure to risk is the amount of time that she was in this state.

When aggregated data are used to calculate rates, exposure to risk is approximated with a midpoint count (or estimate) of the number of persons at risk. Thus, the definitions of fertility rates in Chapter 15 gave counts or estimates of women for the denominators. Individual-level data allow the calculation of exposure to risk on a case-by-case basis.

An *interval of age* has a clear meaning and is generally either a single year or a standard 5-year interval. An *interval of time*, or a *period*, can have two possible interpretations. The first is calendar time. A rate that is calculated for a calendar year, such as 1998, or a 5-year interval, such as 1995–1999, is easily compared with rates from other data sources.

Another possibility is to interpret time as elapsed time before the interview—for example, the year (12 months) prior to the month of interview (perhaps ignoring the actual month of interview, because it has incomplete exposure). An advantage of this interpretation of time is that there is complete exposure for every case, regardless of the date of interview. A disadvantage is that comparisons with other sources are blurred, because the start and end dates of intervals are linked to the dates of fieldwork.

An interval of *marital duration* refers to the length of time since the date of first marriage. Like age, it is generally given in single completed years, beginning with duration zero, or in standard 5-year intervals. For example, marital duration 10 to 14 years begins exactly 10 years after the date of first marriage and ends exactly 15 years after the date of first marriage. It is customary not to re-initialize marital duration if a woman is widowed or divorced or remarries. It is also customary not to make any deductions for time between marriages or to label births between marriages as nonmarital. Such elaborations are possible, if a marital history can be consulted, but are rarely worth the trouble unless there is a specific interest in the measurement of fertility outside of marriage.

The term "window" will refer to an interval of time, stated in century month codes, extending from the beginning to the ending month. A window may be restricted by the requirement that a woman have a specific age or marital duration (in which case it can be thought of as an interval of age or marital duration). A window may be truncated on the left by the date of first marriage, or it may be truncated on the right by the month of interview.

Assume that we know the following for each woman, by calendar month and year:

Her date of birth (needed for the calculation of age)
Her date of first marriage (needed for the calculation of marital duration)
The dates of birth of all her live births
The date of interview

The following kinds of measures are commonly calculated from such data:[2]

Age-period-specific fertility rates
Age-period-specific marital fertility rates
Marital duration-period-specific marital fertility rates
Order-period-specific fertility rates
Period-specific-birth probabilities

All such measures are specific for a time period, typically 5-year intervals of calendar years or of years before the interview month. Our examples will assume 5-year intervals of calendar years, as well as 5-year intervals of age and marital duration. The measures designated as rates in this list fall in the class of "central" rates when calculated from aggregrate data.

To calculate fertility measures from survey data, it is desirable to have a survey that includes all women, not just ever-married women. We shall assume that the data include all women, but will describe the modifications to estimate rates using only ever-married women.

Period Rates

Age-Specific Fertility Rates

We now turn to the most common kinds of specific fertility rates. These rates all have the form of the average or

[2] Conventionally, the names of these measures do not incorporate a reference to the fact that they are period-specific, but this is done here to describe them more fully and to elucidate the method of derivation.

expected number of births to a woman, controlling variously for time period, age, marital status, marital duration, or the order of the birth.

The age-period-specific fertility rate was introduced in the preceding chapter, in which it was calculated with vital statistics data in the numerator and census data (or intercensal or postcensal estimates) on numbers of women in the denominator. Here we calculate it entirely from survey data.

If a refers to an age interval and t to a time interval or period, then the rate for this age and period is calculated as

$$f(a,t) = b(a,t)/e(a,t) \quad (16.1)$$

where $b(a,t)$ is the total number of births observed at time t to women aged a at time of birth and $e(a,t)$ is the total woman-years of exposure to risk at age a during time t. Such rates are sometimes defined to include a factor of 1000, but we shall omit such a factor.

For example, the numerator of the age-period-specific fertility rate for age 20 to 24 in period 1990–1994 consists of the number of births that occurred in 1990–1994 to women who were aged 20 to 24 at the time of the birth. The denominator consists of the total time that the women in the survey were exposed to ages 20 to 24 in 1990–1994. The computational strategy is to examine each woman in the file and locate the window when (if ever) she satisfied the age and period requirements for the rate. The woman's exposure to risk will then be the length of this window. Her relevant births (if any) will be those that occurred within this window. Births and exposures are accumulated for all women in the survey, and then the accumulated births are divided by the accumulated exposures.

We shall illustrate the strategy in detail, continuing with an age-period-specific rate for ages 20 to 24 in 1990–1994. January 1990 converts to cmc = 1 + 90 ∗ 12 = 1081 and December 1994 converts to 12 + 94 ∗ 12 = 1140. (Because December 1994 is 1 month less than 5 years after January 1990, the last month can also be calculated as 1081 + 59 = 1140.) Therefore the window for 1990–1994 is expressed as (1081 to 1140).

The window of time when a woman was aged 20 to 24 must be calculated separately for every woman in the survey. If the cmc of the woman's birth is called B, then she turned 20 in month B + 20 ∗ 12, and 60 months later, she turned 25. The month before this was the final month in her window for ages 20 to 24. Therefore, the window of time when she was aged 20 to 24 is (B + 240 to B + 299).

Most women will have no exposure to ages 20 to 24 during 1990–1994, or to any other specific combination of age and period. A woman will have exposure to this combination of age and period only if her 20th birthday occurred in or before December 1994 *and* her 25th birthday occurred in or after January 1990. For any other woman, the period window (1081 to 1140) and the age window (B + 240 to B + 299) will not intersect. Most women who have any such exposure will have fewer than 60 months; 60 months (5 years) is the maximum possible. (Calculations of exposure will be in months, with subsequent division by 12 to convert to years.)

To repeat, a woman will have some exposure if B + 240 ≤ 1140 *and* 1081 ≤ B + 299. This condition is equivalent to B ≤ 900 *and* 782 ≤ B and can be stated in terms of a range for B as 782 ≤ B ≤ 900. Suppose, for example, that the woman was born in December 1972, so B = 12 + 72 ∗ 12 = 876. Her window for ages 20 to 24 is (1116 to 1175), and the intersection of this variable window for ages 20 to 24 with the fixed window for period 1990–1994 will be (1116 to 1140). This particular woman's exposure to risk is the length of this window, including the first and the last months: 1140 − 1116 + 1 = 25 months. Her contribution to the numerator of this fertility rate will consist of any births that she had from cmc 1116 to cmc 1140, inclusive. The number of such births is obtained by reviewing the dates in the woman's birth history.

It is efficient to determine each woman's contribution to the numerators and denominators of a full array of rates, not just a single rate. Table 16.1 shows the windows of exposure for the woman in the preceding example, assuming that the interview was conducted in April 1998 (cmc 1180). The window for each relevant age interval is given in the last column of the table, and the window for each relevant time period is given in the bottom row of the table. The intersection of the age and period windows is given inside the table.

Now suppose that this woman had had three births, given in her birth history with these dates: August 1990, March 1994, and January 1998. These dates convert to century month codes 8 + 90 ∗ 12 = 1088, 3 + 94 ∗ 12 = 1131, and 1 + 98 ∗ 12 = 1177, respectively. The number of births in the intervals in the age-by-period array that include these dates is incremented by one; this leads to the contributions to the birth array shown in Table 16.2. A zero in a cell indicates that the woman had exposure to that cell, but no births.

Table 16.3 converts the windows of risk into the contributions this woman would make to the cells of the exposure array, expressed in months. For equal intervals of age and period (e.g., 5 years), a specific woman's contribu-

TABLE 16.1 Windows of Age-by-Period Exposure for an Illustrative Woman Born in December 1972 and Interviewed in April 1998

	Period			
Age	**1985–1989**	**1990–1994**	**1995–1999**	**Window for age**
15–19	1056–1080	1081–1115	—	1056–1115
20–24	—	1116–1140	1141–1175	1116–1175
25–29	—	—	1176–1180	1176–1180
Window for period	1021–1080	1081–1140	1141–1180	

Source: See text for explanation.

tion to the arrays of exposures (and births) will always be located on two adjacent diagonals going from the upper left to the lower right. We give half a month of exposure to the month of interview (and include any births reported in that month).

Three rows and three columns are shown for this respondent's age-by-period array, but they are extracted from a larger array that would extend to ages 45 to 49 and to periods going back as far as desired, up to 35 years (if the age range of eligible respondents is 50 − 15 = 35 years). In practice this kind of array is often taken back no more than 10 years. The further back it goes, the greater the chance of reporting error. If maternal mortality is high, there is omission of the higher-fertility women who have died, along with their births, so that the more remote rates are biased downward. A comparison of the more remote rates with those obtained from an earlier survey can help to identify patterns of reporting error and selectivity.

TABLE 16.2 Contributions of the Illustrative Woman to the Numerators of Age-Period-Specific Fertility Rates

	Period		
Age	1985–1989	1990–1994	1995–1999
15–19	0	1	—
20–24	—	1	0
25–29	—	—	1

Source: See text for explanation.

TABLE 16.3 Months of Age-by-Period Exposure for the Illustrative Woman

	Period		
Age	1985–1989	1990–1994	1995–1999
15–19	25 months	35 months	0 months
20–24	0 months	25 months	35 months
25–29	0 months	0 months	4.5 months

Source: See text for explanation.

These steps are repeated for every woman in the survey, until a full array of exposures and a corresponding full array of rates are obtained. Exposures are converted to years by dividing the months of exposure by 12, and the total births are divided by total exposure, cell by cell, to give the age-period specific fertility rates.

Because there is a cutoff age in a fertility survey (usually age 49), the array of rates will be empty in the lower left of the table. If we go back more than five years, we have no information about women aged 45 and over; if we go back more than 10 years, we have no information about women aged 40 and over; and so on.

Tables 16.4 through 16.6 give the numbers of births, months of exposure, and age-period-specific fertility rates for the 1998 NDHS of the Philippines. The last time interval is labeled "1995–1999", but the data for that interval should be understood to extend only to the fieldwork in 1998. In these and later tables, cells with zeros or dashes should be interpreted to be outside the time and age (or duration) range of the survey.

How would these rates be estimated if the survey was limited to ever-married women? The answer is relatively simple if there is an accompanying household roster that indicates which women were selected for the interviews, and if there is a negligible amount of childbearing by never-married women. Assume that these two conditions are true.

The birth array would be calculated exactly as shown earlier, but would necessarily be limited to the birth histories of the ever-married women. The exposure array, however, would be calculated from all the women in the household roster who were in the age range of the eligible respondents at the time of the survey, usually 15 to 49, and would not be limited to the ever-married women in the survey. The rates would again be calculated by dividing the birth array by the exposure array, cell by cell.

If the researcher wishes to calculate age-period-specific rates within socioeconomic categories, using a survey

TABLE 16.4 Numerators of Age-Period-Specific Fertility Rates (numbers of births): Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
15–19	309	572	732	765	745	710	463
20–24	75	1104	1709	2101	2231	2186	1342
25–29	0	82	1338	1791	2173	2370	1495
30–34	0	0	107	1062	1395	1754	1098
35–39	0	0	0	67	689	882	676
40–44	0	0	0	0	32	315	211
45–49	0	0	0	0	0	5	24

Source: Based on the 1998 National Demographic and Health Survey (NDHS) of the Philippines (Philippines National Statistics Office, Department of Health, and Macro International, 1999).

TABLE 16.5 Denominators of Age-Period-Specific Fertility Rates (years of exposure): Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
15–19	4,789.58	7,184.83	9,040.25	10,204.92	10,753.75	11,145.50	8,717.04
20–24	364.58	4,789.58	7,184.83	9,040.25	10,204.92	10,753.75	7,091.21
25–29	0	364.58	4,789.58	7,184.83	9,040.25	10,204.92	6,926.50
30–34	0	0	364.58	4,789.58	7,184.83	9,040.25	6,589.42
35–39	0	0	0	364.58	4,789.58	7,184.83	5,646.67
40–44	0	0	0	0	364.58	4,789.58	4,469.67
45–49	0	0	0	0	0	364.58	2,626.13

Source: Based on the 1998 NDHS of the Philippines (Philippines National Statistics Office, Department of Health, and Macro International, 1999).

TABLE 16.6 Age-Period-Specific Fertility Rates (births per year); Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
15–19	.0645	.0796	.0810	.0750	.0693	.0637	.0531
20–24	.2057	.2305	.2379	.2324	.2186	.2033	.1892
25–29	—	.2249	.2794	.2493	.2404	.2322	.2158
30–34	—	—	.2935	.2217	.1942	.1940	.1666
35–39	—	—	—	.1838	.1439	.1228	.1197
40–44	—	—	—	—	.0878	.0658	.0472
45–49	—	—	—	—	—	.0137	.0091

Source: Tables 16.4 and 16.5.

limited to ever-married women, this will be possible only for categories specified for the women in the household survey. Region, type of residence, and education are usually available in the household survey, but not much more. In an all-women survey, rates can be made specific for any characteristics of the respondents. Rates that are defined only for ever-married women, such as the marital fertility rates to be discussed next, are obviously unaffected by the restriction of eligibility to ever-married women.

Age-Specific Marital Fertility Rates

Marital fertility rates are restricted according to marital status, but otherwise they are calculated in the same way as rates that do not refer to marital status. The marital fertility rate for age *a* and period *t* is

$$mf(a,t) = mb(a,t)/me(a,t) \qquad (16.2)$$

where $mb(a,t)$ is the total marital births observed at time *t* to women aged *a* at time of birth and $me(a,t)$ is the total woman-years of marital exposure to risk at age *a* at time *t*.

This type of rate is limited to births and exposure that occur after first marriage. As stated earlier, the marriage history is typically not consulted for any dates other than the date of first marriage (or when the couple first lived together as "married"). If M is the month of first marriage, then any window will omit exposure and births prior to M. For example, if the respondent in the previous example was married in February 1993 (i.e., $M = 2 + 93 * 12 = 1118$), then the window (1116–1140) will be reduced to (1118–1140); this will be the respondent's window of exposure to marital fertility while at ages 20 to 24 in period 1990–1994; the exposure to risk will be $1140 - 1118 + 1 = 23$ months. Births prior to month 1118 will be ignored, leaving two marital births, in months 1131 and 1171. After the window in which the first marriage occurred, there will be no difference between a woman's contributions to the numerator and denominator of the marital rate and the overall rate. The windows of risk, marital births, and marital exposure for the illustrative woman are given in Tables 16.7 through 16.9.

TABLE 16.7 Windows of Age-by-Period Marital Exposure for the Illustrative Woman

	Period		
Age	1990–1994	1995–1999	Window for Age
15–19	—	—	1056–1115
20–24	1118–1140	1141–1175	1116–1175
25–29	—	1176–1180	1176–1180
Window for period	1081–1140	1141–1180	

Source: See text for explanation.

TABLE 16.8 Marital Births for the Illustrative Woman

	Period	
Age	1990–1994	1995–1999
15–19	0	—
20–24	1	0
25–29	—	1

Source: See text for explanation.

TABLE 16.9 Months of Age-by-Period Marital Exposure for the Illustrative Woman

	Period	
Age	1990–1994	1995–1999
15–19	0 months	0 months
20–24	23 months	35 months
25–29	0 months	4.5 months

Source: See text for explanation.

TABLE 16.10 Numerators of Marital-Age-Period-Specific Fertility Rates (births): Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
15–19	299	544	698	727	707	681	432
20–24	73	1081	1681	2069	2194	2142	1312
25–29	0	79	1317	1782	2155	2342	1481
30–34	0	0	105	1058	1391	1749	1097
35–39	0	0	0	67	685	880	674
40–44	0	0	0	0	31	315	211
45–49	0	0	0	0	0	5	24

Source: Based on 1998 NDHS of the Philippines (Philippines National Statistics Office, Department of Health, and Macro International, 1999).

TABLE 16.11 Denominators of Marital-Age-Period-Specific Fertility Rates (years of exposure): Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
15–19	747.75	1374.92	1655.92	1756.83	1710.50	1625.83	965.38
20–24	153.00	2601.42	4229.58	5170.42	5625.50	5649.92	3511.42
25–29	0	253.58	3809.92	5945.92	7223.08	8134.33	5472.13
30–34	0	0	314.58	4316.83	6498.42	8101.58	5954.83
35–39	0	0	0	329.75	4446.33	6682.25	5250.63
40–44	0	0	0	0	336.42	4498.42	4240.17
45–49	0	0	0	0	0	338	2507.29

Source: Based on 1998 NDHS of the Philippines (Philippines National Statistics Office, Department of Health, and Macro International, 1999).

Working from a data file, each respondent's contributions to the arrays for marital births and marital exposures are calculated; months of marital exposure are converted to years; and the rates are calculated by dividing births by exposure, cell by cell. Tables 16.10 through 16.12 give the numbers of marital births, months of marital exposure, and age-period-specific marital fertility rates for the 1998 NDHS of the Philippines.

Marital Duration-Specific Marital Fertility Rates

Marital fertility can be referenced by the woman's date of marriage rather than her date of birth. The marital fertility rate for duration d and period t is

$$mf(d,t) = mb(d,t) / me(d,t) \qquad (16.3)$$

where $mb(d,t)$ is the total marital births observed at time t to women with duration d at time of birth and $me(d,t)$ is the total woman-years of marital exposure to risk at duration d at time t.

For example, if the illustrative woman was married in month M = 1118, then she had marital duration 0 to 4 years in the window (M to M + 59)—that is, (1118 to 1177), and so on. Her windows of exposure, marital births, and months of exposure for the marital duration rates are given in Tables 16.13 through 16.15.

The marital duration-period-specific rates are calculated, as shown earlier, by dividing the accumulated array of births by the accumulated array of exposures (converted to years). Tables 16.16 through 16.18 give the numbers of marital births, the months of marital exposure, and marital duration-

TABLE 16.12 Marital-Age-Period-Specific Fertility Rates (births per year): Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
15–19	.3999	.3957	.4215	.4138	.4133	.4189	.4475
20–24	.4771	.4155	.3974	.4002	.3900	.3791	.3736
25–29	—	.3115	.3457	.2997	.2983	.2879	.2706
30–34	—	—	.3338	.2451	.2141	.2159	.1842
35–39	—	—	—	.2032	.1541	.1317	.1284
40–44	—	—	—	—	.0921	.0700	.0498
45–49	—	—	—	—	—	.0148	.0096

Source: Tables 16.10 and 16.11.

TABLE 16.13 Windows of Marital Duration-by-Period Exposure for the Illustrative Woman

		Period		Window for duration
		1990–1994	1995–1999	
Marital	0–4	1118–1140	1141–1177	1118–1177
duration	5–9	—	1178–1180	1178–1180
Window for period			1081–1140	1141–1180

Source: See text for explanation.

TABLE 16.14 Marital Births for the Illustrative Woman

		Period	
		1990–1994	1995–1999
Marital	0–4	1	1
duration	5–9	—	0

Source: See text for explanation.

TABLE 16.15 Months of Marital Duration-by-Period Exposure for the Illustrative Woman

		Period	
		1990–1994	1995–1999
Marital	0–4	23 months	37 months
duration	5–9	0 months	2.5 months

Source: See text for explanation.

period-specific marital fertility rates for the 1998 NDHS of the Philippines.

Age-Period-Cohort Relationships

Each of the arrays of fertility rates described can be examined from any of three perspectives or directions. One of these perspectives is the woman's life course, indicated by her age (in the case of age-period-specific fertility rates and age-period-specific marital fertility rates) or by her marital duration (in the case of duration-period-specific marital fertility rates), as shown in the columns of Tables 16.6, 16.12, and 16.18. The second perspective is across time periods, as shown in the rows of these tables.

A third perspective, perhaps less obvious, is across birth cohorts (in the case of the first two kinds of rates) or across marriage cohorts (in the case of the third kind of rates), as shown by the diagonals extending from the upper left to the lower right of these tables. Recall that a birth cohort consists of persons born in the same time interval and a marriage cohort consists of persons married in the same time interval.

In an array of age-period-specific rates, for example, rates in the same row, referring to the same age interval, can be compared across columns or periods, to identify patterns of change over time. They can also be compared across diagonals, to identify patterns of change across birth cohorts.

Rates calculated for the typical 5-year interval of age/duration and 5-year interval of time will draw from a 10-year (rather than 5-year) cohort of births. For example, women who are aged 25 to 29 in any part of the time interval 1995–1999 could have been born as early as 1995 – 30 = 1965 and as late as 1999 – 25 = 1974 (i.e., anytime during the 10 years between January 1, 1965 and December 31, 1974). This feature of the widths of intervals carries over to a larger class of rates. We have not described age-cohort-specific rates, for example, although the procedures for calculating them are very similar to those described for age-period-specific rates. If such rates were calculated for 5-year age groups and 5-year birth cohorts, then the period intervals would be spread over 10 years. In general, if the intervals for the first two dimensions are $w1$ and $w2$ years, then an interval for the third dimension will be $w3 = w1 + w2$ years. The wider intervals will overlap one another. For

TABLE 16.16 Numerators of Marital Duration-Period-Specific Fertility Rates (births): Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
0–4	348	1372	2374	3065	3346	3511	2334
5–9	24	319	1126	1649	1982	2148	1292
10–14	0	13	283	783	1170	1306	808
15–19	0	0	18	187	530	766	505
20–24	0	0	0	19	130	346	235
25–29	0	0	0	0	5	36	56
30–34	0	0	0	0	0	1	1

Source: Based on 1998 NDHS of the Philippines (Philippines National Statistics Office, Department of Health, and Macro International, 1999).

TABLE 16.17 Denominators of Marital Duration-Period-Specific Fertility Rates (years of exposure): Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
0–4	838.17	3272.25	5700.58	7432.00	8218.42	9116.92	6200.21
5–9	62.58	895.08	3347.92	5778.33	7531.58	8290.67	5862.58
10–14	0	62.58	898.92	3347.92	5780.83	7532.50	5260.29
15–19	0	0	62.58	898.92	3347.92	5780.83	4802.75
20–24	0	0	0	62.58	898.92	3347.92	3500.96
25–29	0	0	0	0	62.58	898.92	1870.42
30–34	0	0	0	0	0	62.58	389.88

Source: Based on 1998 NDHS of the Philippines (Philippines National Statistics Office, Department of Health, and Macro International, 1999).

TABLE 16.18 Marital Duration-Period-Specific Fertility Rates (births per year): Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
0–4	.4152	.4193	.4164	.4124	.4071	.3851	.3764
5–9	.3835	.3564	.3363	.2854	.2632	.2591	.2204
10–14	—	.2077	.3148	.2339	.2024	.1734	.1536
15–19	—	—	.2876	.2080	.1583	.1325	.1051
20–24	—	—	—	.3036	.1446	.1033	.0671
25–29	—	—	—	—	.0799	.0400	.0299
30–34	—	—	—	—	—	.0160	.0026

Source: Tables 16.16 and 16.17.

example, cohorts described with the diagonals of the usual age-period-specific rates will refer to birth dates such as 1950–1959, 1955–1964, 1960–1969, 1965–1974, and so on. This blurring in the third dimension, so to speak, has the effect of suppressing some of the variation in that dimension in the same way that a moving average does. The linkages between age/duration, period, and cohort are relevant to all demographic and sociocenomic variables that have an age dimension (e.g., mortality, labor force).

Order-Specific Fertility Rates

As may be recalled from Chapter 15, a woman's parity is the number of live births that she has had. Any of the preceding rates can also be made specific for the parity of the mother or, equivalently, the birth order of the latest child. Such rates are sometimes called parity-specific, with reference to the mother, and sometimes called order-specific, with reference to the child; we shall describe them as order-specific. These rates were introduced in Chapter 15, but their construction will be briefly reviewed in the present context.

It is easiest to clarify the labeling with an example. An age-period-order-specific rate is specific for order two if it measures the rate of childbearing of second births for women in each combination of age and period. When a second birth occurs, the woman moves from parity one to parity two. Thus the rate is indexed by the birth that closes the interval.

By convention, the denominators of these rates *do not* depend on the woman's parity. They do not reflect the fact that, say, a woman who has had no births at all has no immediate risk of having a second birth, or that, say, a woman who has had one or more births is no longer at risk of having a first birth. Only the woman's contribution to the numerator depends on her parity. That contribution will be one birth if she has a birth of the specified order in the age-period window. Thus, the numerators of the three-way rates are formed by disaggregating the numerators of the two-way rates. As a result, when the age-period-order-specific rates are added up across all birth orders, the sum will be simply the age-period-specific rate.

The order-specific rate for births of order *j*, for age *a* at time *t*, is given by

$$f_j(a,t) = b_j(a,t)/e(a,t) \tag{16.4}$$

where $b_j(a,t)$ is the number of births *of order j* and $f(a,t) = \sum_j f_j(a,t)$.

Consider again the illustrative woman who was born in December 1972 and had births in century months 1088, 1131, and 1177. This respondent's windows of age and period were given earlier, and her months of exposure to risk were given in Table 16.3. These will be her contributions to the denominator of *every* order-specific rate within the combinations of age and period. We now wish to identify her contributions to fertility rates that are specific for age, period, and order.

The woman's three births were classified by age and period in Table 16.2. We now repeat that table, with "first," "second," and "third" inserted in the table to identify birth order, as shown in Table 16.19.

The first birth will contribute only to the numerator of the age-period-order-specific rate for ages 15 to 19, period 1990–1994, order 1. The second birth will contribute only to the numerator of the age-period-order specific rate for ages 20 to 24, period 1990–1994, order 2. The third birth will contribute only to the numerator of the age-period-order-specific rate for ages 25 to 29, period 1995–1995, order 3. The woman contributes nothing to the numerators of any other age-period-order-specific rates.

It would be possible also to disaggregate the woman's exposure to risk in order to describe the months of exposure to risk of a first birth, second birth, and so on, within each cell of the exposure (denominator) array and to calculate order-specific rates that controlled for parity in the same way as for age and period, but we emphasize that the conventional order-specific rates do not do this.

If desired, any of the other rates that are specific for possible combinations, i.e., of age, period, cohort, marital status, or marital duration, can also be made specific for birth order.

Tables 16.20 and 16.21 give the arrays of births by age, period, and birth order, for birth orders one and two, from the 1998 NDHS of the Philippines. Tables 16.22 and 16.23 give the corresponding age-period-order-specific rates, obtained by dividing the successive panels of births by the exposures in Table 16.5.

Birth Probabilities

A fertility rate is essentially an average or expected *number of births* that occur in an interval. A birth probability, by contrast, is the (estimated) *probability that one or more births* will occur within that interval. A retrospective

TABLE 16.19 Contributions of the Illustrative Woman to the Numerators of Order-Age-Period-Specific Fertility Rates

	Period		
Age	**1985–1989**	**1990–1994**	**1995–1999**
15–19	0	1 (first)	—
20–24	—	1 (second)	0
25–29	—	—	1 (third)

Source: See text for explanation.

TABLE 16.20 Numerators of Order-Age-Period-Specific Fertility Rates for Order 1: Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
15–19	208	399	478	519	509	482	336
20–24	32	389	546	696	791	788	543
25–29	0	14	195	246	321	392	289
30–34	0	0	12	73	86	123	90
35–39	0	0	0	2	12	31	32
40–44	0	0	0	0	3	9	2
45–49	0	0	0	0	0	0	0

Source: Based on 1998 NDHS of the Philippines (Philippines National Statistics Office, Department of Health, and Macro International, 1999).

TABLE 16.21 Numerators of Order-Age-Period-Specific Fertility Rates for Order 2: Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
15–19	79	145	191	195	183	176	107
20–24	30	354	563	678	730	713	436
25–29	0	11	241	363	422	499	326
30–34	0	0	15	115	133	196	150
35–39	0	0	0	6	34	34	39
40–44	0	0	0	0	0	6	5
45–49	0	0	0	0	0	0	1

Source: Based on 1998 NDHS of the Philippines (Philippines National Statistics Office, Department of Health, and Macro International, 1999).

TABLE 16.22 Order-Age-Period-Specific Fertility Rates for Order 1: Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
15–19	.0434	.0555	.0529	.0509	.0473	.0432	.0385
20–24	.0878	.0812	.0760	.0770	.0775	.0733	.0766
25–29	—	.0384	.0407	.0342	.0355	.0384	.0417
30–34	—	—	.0329	.0152	.0120	.0136	.0137
35–39	—	—	—	.0055	.0025	.0043	.0057
40–44	—	—	—	—	.0082	.0019	.0004
45–49	—	—	—	—	—	—	.0000

Source: Tables 16.20 and 16.5.

TABLE 16.23 Order-Age-Period-Specific Fertility Rates for Order 2: Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
15–19	.0165	.0202	.0211	.0191	.0170	.0158	.0123
20–24	.0823	.0739	.0784	.0750	.0715	.0663	.0615
25–29	—	.0302	.0503	.0505	.0467	.0489	.0471
30–34	—	—	.0411	.0240	.0185	.0217	.0228
35–39	—	—	—	.0165	.0071	.0047	.0069
40–44	—	—	—	—	—	.0013	.0011
45–49	—	—	—	—	—	—	.0004

Source: Tables 16.21 and 16.5.

survey is actually the only data format (other than a prospective survey, which is only rarely available) that allows the direct calculation of birth probabilities. We will briefly show how this may be done, staying with 5-year intervals of age and time.

In the context of rates, we described the construction of a numerator array of births and a denominator array of exposures. Each woman made contributions (often zero) to the cells of the numerator and denominator arrays. For each woman, each cell was expressed in terms of a window of century months, within which births and exposure may have occurred. The rates were subsequently calculated from a cell-by-cell division of the accumulated contributions to the numerator and denominator arrays. In a typical situation where exposure is calculated in months but we want a single-year rate, at some point (most easily after all observations have been accumulated) the exposure must be divided by 12.

The approach for probabilities is similar, but the arrays of births and exposures are calculated slightly differently. For probabilities, an individual woman's contribution to a numerator cell of births can only be zero if she had no births in the window, or one if she had one *or more* births in the window. An algorithm for calculating her contribution to the numerator of a rate needs only to be altered by recoding any positive number of births into just one birth. A woman's contribution to the corresponding denominator cell of exposure will depend in part on whether she actually had a birth during the window of observation. This may seem counterintuitive and calls for some justification.

Consider the probability that a woman will have a birth while aged 20 to 24, which is to be estimated with births and exposure observed within that age interval during 1990–1994. The outcome is binary: Either a birth occurs, in which case a code of 1 is assigned, or no birth occurs, in which case a code of 0 is assigned.

The classical example of a trial with a binary outcome is the toss of a coin. We toss a coin once, and assign code 1 to a success, say a head, and code 0 to a failure, a tail. The numerator or outcome can be 0 or 1, interpreted as the (possible) number of heads or successes when the number of trials—the denominator—is 1. If we tossed n independent but identical coins, the number of successes could be any integer k between 0 and n, inclusive, and k would have a binomial distribution with denominator n.

This familiar model may clarify the requirement that if the outcome takes the values 0 or 1, then the denominator, or degree of risk associated with the outcome, must never be less than the value of the numerator. If the denominator could be less than 1 when the numerator is 1, we would have the potential to produce an estimated probability greater than 1, which is not allowed.

The relevant data in a window of age and time are generally censored on either the left or the right, and sometimes (if the interview occurred within the time interval) on both the left and the right. There are three possibilities.

If the observation is not censored, and the woman had a full 60 months of exposure to the window, then her contribution to the denominator will be 1 (the number of months in the window divided by 60).

If the observation is censored, *and no birth occurred in the window*, then her contribution to the denominator will be the *fraction* of the full 5-year or 60-month interval for which she was observed—that is, the number of months in the window divided by 60. This fraction indicates that the observation is only partial.

The remaining possibility is that the observation is censored but a birth does occur. This is where the coin-tossing analogy becomes relevant. *If a birth occurred*, then *it does not matter* that there was less than full observation of the woman, and we credit the case with a contribution of 1 to the denominator. Indeed, we *must* credit her with 1 to avoid having a contribution to the denominator that is less than the contribution to the numerator.

To summarize,

- If there was *no censoring* within the cell, then exposure for the probability equals the exposure for the rate.
- If there was *censoring* and *no birth*, then exposure for the probability equals the exposure for the rate.
- If there was *censoring* and *a birth* (one or more), then exposure must be augmented to reach the length it would have had in the absence of censoring (e.g., 60 months).
- After the exposure in a cell has been accumulated across all respondents, the sum must be normalized to a maximum of one unit per woman (e.g., by dividing the total months by 60).

These rules are consistent with an assumption that the probability of having a birth is uniform within the interval of age and time (or intervals of other dimensions, depending on the specific rate). If this assumption is not plausible, then (as with a rate) the researcher may choose to adopt shorter intervals, within which the assumption is safer.

Reconsider the illustrative woman born in December 1972, with births in century-months 1088, 1131, and 1177. We described in detail the calculation of this woman's contributions to the numerator and denominator arrays of age-period-specific fertility rates. How would these contributions differ for age-period-specific birth probabilities?

First consider the numerator array. Cells with no births will continue to make a contribution of zero. Because the woman's births occurred in different cells of the age-by-period array, the three cells with a contribution of one birth to the numerator of a rate also contribute one birth to the numerator of a probability. The numerator array will thus be exactly the same as Table 16.3 and need not be repeated.

The contributions to the denominator array will remain unchanged for those cells in which no births occurred. In the three cells in which a birth occurred, the months of exposure must be increased to 60. After the accumulation of all exposures, we emphasize that the total in each cell must be normalized by dividing by 60 months, rather than 12 months, and that the probability extends across a 5-year range, whereas the rate is interpreted in terms of a single year. The denominator contributions are given in Table 16.24, prior to the division by 60.

Tables 16.25 through 16.27 give the births, exposures, and age-period-specific birth probabilities for the 1998

TABLE 16.24 The Illustrative Woman's Contributions to Risk for the Age-Period-Specific Birth Probabilities

	Period		
Age	**1985–1989**	**1990–1994**	**1995–1999**
15–19	25 months	60 months	0 months
20–24	0 months	60 months	35 months
25–29	0 months	0 months	60 months

Source: See text for explanation.

TABLE 16.25 Numerators of Age-Period-Specific Birth Probabilities (births): Philippines: 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
15–19	391	450	568	599	588	558	401
20–24	72	732	1213	1458	1613	1566	1097
25–29	0	79	920	1327	1622	1785	1255
30–34	0	0	104	777	1079	1336	945
35–39	0	0	0	65	536	715	587
40–44	0	0	0	0	32	276	197
45–49	0	0	0	0	0	5	24

Source: Based on 1998 NDHS of the Philippines (Philippines National Statistics Office, Department of Health, and Macro International, 1999).

TABLE 16.26 Denominators of Age-Period-Specific Birth Probabilities (years of risk): Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
15–19	5,041.75	7,989.92	10,129.08	11,276.92	11,820.58	12,157.92	9,723.63
20–24	643.42	5,932.33	9,438.33	11,715.92	13,163.58	13,546.17	9,868.00
25–29	0	670.50	6,237.83	9,693.83	12,006.67	13,511.42	10,091.54
30–34	0	0	773.33	6,021.58	9,189.33	11,454.58	8,929.33
35–39	0	0	0	613.00	5,616.83	8,521.17	7,096.38
40–44	0	0	0	0	486.67	5,224.08	4,968.38
45–49	0	0	0	0	0	383.25	2,694.38

Source: Based on 1998 NDHS of the Philippines (Philippines National Statistics Office, Department of Health, and Macro International, 1999).

TABLE 16.27 Age-Period-Specific Birth Probabilities (probability of a birth in 5 years): Philippines, 1998

Age	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
15–19	.3878	.2816	.2804	.2656	.2487	.2295	.2062
20–24	.5595	.6170	.6426	.6222	.6127	.5780	.5558
25–29	—	.5891	.7374	.6845	.6755	.6606	.6218
30–34	—	—	.6724	.6452	.5871	.5832	.5292
35–39	—	—	—	.5302	.4771	.4195	.4136
40–44	—	—	—	—	.3288	.2642	.1983
45–49	—	—	—	—	—	.0652	.0445

Source: Tables 16.25 and 16.26. Calculation: (Cell in Table 16.25 ÷ Cell in Table 16.25) × 5.

NDHS of the Philippines. Because these probabilities refer to 5-year intervals of age, they are much larger than the corresponding annual rates given in Table 16.6.

Standard Errors and Sample Design

Standard Errors

When sample data are used to calculate fertility rates or birth probabilities, it should be clearly understood that these are *estimates* of the rates and probabilities for the population from which the sample is drawn. We have followed demographic practice in referring to the preceding quantities as birth probabilities, but they are actually only estimates of birth probabilities. They are subject to sampling error, therefore, measured in terms of standard errors. A standard error can be interpreted as the average deviation (ignoring direction) of an estimate from the true (population) value across all possible random samples of the same size. It can be used for constructing interval estimates (confidence intervals) and for testing hypotheses about the rates and probabilities in the population. Fortunately, it is fairly easy to produce good estimates of the standard errors of specific rates.

The generic form for a specific rate is $r = b/e$, where r is the rate (ignoring any multipliers such as 1000), b is a count or frequency of births, and e is a measure of exposure in woman-years. Let $s.e.(r)$ denote the estimated standard error of the rate.

As a good first approximation, for a fixed amount of exposure, the number of births has a Poisson distribution. A useful property of a Poisson distribution is that its mean and variance are equal. The observed number of births will be the maximum likelihood estimate of both the mean and the variance of the distribution. Therefore,

$$s.e.(r) = \sqrt{b}\big/e = \sqrt{r/e} = r\big/\sqrt{b}. \qquad (16.5)$$

Any of the three forms on the right-hand side of this equation can be used. Suppose, for example, that an age-specific rate for ages 30 to 34 is .100 births per woman per year, and the numerator of this rate included 400 births. Then the estimated standard error of the rate would be

$$s.e.(r) = r\big/\sqrt{b} = .100\big/\sqrt{400} = .005$$

Another useful property of a Poisson sampling distribution for a birth count is that it is well approximated by a normal distribution having the same mean and standard deviation, especially for large samples. Adapting the usual formulas for confidence intervals for parameters whose estimates have asymptotically normal sampling distributions, a 95% confidence interval for the underlying true rate will be $r \pm 1.96r/\sqrt{b}$. The 95% confidence interval for the rate estimated in the preceding paragraph would thus range from .090 to .110.

Suppose there are two independent estimates r_1 and r_2 of the true rates for two subpopulations (or two time periods, two age groups, etc.). Then the test statistic for a null hypothesis that the underlying rates are equal will be

$$z = (r_1 - r_2)\big/\sqrt{(r_1^2/b_1) + (r_2^2/b_2)}. \qquad (16.6)$$

For example, a two-sided null hypothesis will be rejected at the .05 level if the calculated test statistic is greater than 1.96 or less than −1.96.

The standard error for an estimated birth probability is estimated by drawing on statistical theory for a binomial distribution. Say that the estimated probability is $p = b/e$, where both b and e are different from the preceding discussion of rates; here b is the total relevant birth count limited to 0's and 1's, and e is the total exposure, scaled to be 1 if a birth occurred or there was no censoring, or the appropriate fraction if no birth occurred and there was censoring. Then the standard error of p is estimated to be

$$s.e.(p) = \sqrt{p(1-p)/e} = \left(p\big/\sqrt{b}\right)\left(\sqrt{1-p}\right). \qquad (16.7)$$

The standard error of p is similar to the standard error of r, particularly when p (or r) is small. Formulas for confidence intervals and test statistics using estimated probabilities and standard errors can be found in statistics texts.

Sample Design

In a simple random sample, every case in the population has the same probability of appearing in the sample and this probability is independent of whether another case appears. Statistical theory is based on such a model, but virtually no fertility survey follows these criteria. Most surveys have a stratified cluster design, in which relatively small subpopulations are oversampled and relatively large subpopulations are undersampled. Census enumeration districts or other such administrative areas comprise the primary sampling units, within which households and individuals are selected. These departures from the model of a simple random sample have two important implications.

First there is the issue of sampling weights, which compensate for oversampling and undersampling. Such weights are inversely proportional to the probability that a case in the sample would have been selected from the population. If a case was oversampled, for example, the weight would be relatively small, and if it was undersampled, the weight would be relatively large. If needed, weights are calculated by the survey organization and included on each computer record, generally near the case identification codes. They are generally constructed so that (if the decimal point is properly located) the average value of the weights is 1.0, and the total of the weights, across the entire sample, is equal to the number of cases in the sample.

If weights are provided, we recommend that they be used in the calculation of the descriptive measures given in this chapter (even though we did not follow that practice for the illustrations with data from the Philippines). Otherwise, the measures will be biased toward the oversampled subpopulations. The weights compensate for this bias. Statistical computer packages generally have a weight option. The researcher simply invokes that option and identifies the weight variable. If x_i is the value of a variable for case i, and w_i is the weight for case i, then the main effect of the weight option is to replace x_i by $w_i x_i$, to replace x_i^2 by $w_i x_i^2$, and so on. In the calculation of mean CEB, for example, if x_i is the CEB for woman i, the mean CEB would be $\sum_i w_i x_i \big/ \sum_i w_i$. The weight w_i appears in the denominator in place of an implied count of 1 for case i. An unweighted estimate has a slightly smaller standard error than a weighted estimate—which is one reason why some researchers do not use weights.

There is uniform agreement that *for the calculation of standard errors*, every real case in the sample should be given equal importance. This practice means that parallel computer runs are often required: one with weights for estimation and one without weights to get standard errors for the construction of confidence intervals or tests of hypotheses.

Thus, the recommended practice with respect to weights is as follows:

- Use sampling weights when unbiased descriptive estimates are desired.
- Use sampling weights in multivariate models unless the model includes all the stratifying variables or you are confident that the model is fully specified.
- Omit sampling weights for the estimation of standard errors *or* use software that gives "robust" weighted estimates.

The second issue for the analysis of complex sampling designs comes from the use of sample clusters. The cases in these clusters are not independent, but overlap in the information they provide. The degree of overlap is reflected in the intraclass correlation. This is not easily calculated and varies from one variable to another. The lack of independence will not alter the estimates of summary statistics, rates, probabilities, and coefficients, but it will affect standard errors. If the clustering is not taken into account, the estimated standard errors will tend to be too small, and as a result, the *p*-values in hypothesis tests will be understated and confidence intervals will be too narrow. If possible, the magnitude of these effects should be assessed with a computer package.

MEASURES OF OVERALL FERTILITY

Overall fertility refers to the total number of births relative to the total number of persons or women in the population. Age and marital duration are ignored, except where women are restricted to the range of the childbearing ages. These measures are described after the specific rates, rather than before them, because they involve some concepts introduced in connection with the specific rates.

Crude Birthrate

The crude birthrate (CBR) was defined in Chapter 15 to be the number of births in a fixed reference period, generally a year, divided by the (total) population at the midpoint of the reference period, multiplied by 1000. Typically, the numerator comes from vital statistics and the denominator comes from a census or is estimated from census data.

It is possible to estimate the CBR entirely from survey data if the survey includes a roster for all the sample households, including the households that had no eligible respondents. The household sample then represents the general population at the time of data collection. The birth histories provide a count of the number of births during a recent time interval for the sample, giving the numerator of the CBR. The total household count can serve as the denominator of the CBR for recent time periods. Because of the general interest in the CBR, we will go into some detail on the issues raised when it is estimated with such data.

The fieldwork for the 1998 NDHS of the Philippines was conducted entirely in 1998, so the survey could be used to estimate the CBR in 1997. The numerator would consist of all births observed in the birth histories for 1997, namely 1586 births (Philippines NDHS Report, 1999, Table C4). An approximation for the denominator would be the total number of persons in the household survey, namely 60,349 persons (Philippines NDHS Report, 1999, Table C1). The ratio, multiplied by 1000, is $(1586/60{,}349) \times 1000 = 26.3$. In this illustration, the numerator and denominator are unweighted. The weighted estimate of the CBR for the 36 months before the survey is 28.0 (Philippines NDHS Report, 1999, Table 3.1).

The standard error of the unweighted estimate of the CBR for 1997 is $(\sqrt{1586}/60{,}349) \times 1000 = .66$, so that a 95% confidence interval for the estimated rate (unweighted) would range from 25.0 to 27.6 ($= 26.3 \pm 1.96$ times .66). Following the practice of ignoring weights for the calculation of standard errors, we could also use .66 to construct a confidence interval for the *weighted* CBR, so that a 95% confidence interval for the weighted estimate would range from 26.7 to 29.3 ($= 28.0 \pm 1.96$ times .66).

The median date of the 1998 NDHS fieldwork was approximately April 1, 1998; the midpoint of 1997 was July 1, 1997. One might argue that a denominator at the time of the survey is too large and could be improved by projecting the household population back nine months, or .75 of a year, using the Philippines' estimated annual growth rate of 2.0%. We would not advocate such an adjustment because it ignores an inherent linkage between the numerator and denominator data. If we deflated the denominator population, then to be consistent we should also deflate the number of women who produced the births, and this in turn would deflate the number of births. The same adjustments would be made to both the numerator and the denominator and they would cancel out (assuming no change in the birth rate in this period).

Another way to improve the denominator would be to use the household survey to calculate person-years lived by all household members during 1997. Although this step would be an improvement, such a denominator would ignore anyone who had died between the beginning of 1997 and the date of the survey, and would be somewhat too *small*. Whatever adjustments to the denominator one might make for a recent time interval, they are unlikely to be outside the range of sampling error.

Compared with the traditional definition, the numerator described earlier omits births to women who had a birth during the reference window of time but died between the birth and the interview. Such women and their births are omitted from the survey. This effect is small unless adult female mortality is extremely high. The numerator also omits births to women who were near the end of the eligible age range at the beginning of the window and "aged out" by the date of interview (e.g. who turned 50 between these two dates). This effect is also negligible (unless the window is backdated several years before the interviews) because women near the upper end of eligibility have very low fertility. Because of the biases in the numerator and denominator, it is safest to limit the survey estimate of the CBR to a recent time interval, but in order to gain statistical stability, an interval longer than 1 year is desirable. DHS reports typically include an estimate of the CBR for 1 to 36 months before the interview. A 3-year estimate will have a smaller standard error, by a factor of approximately $1/\sqrt{3} = .58$.

Child-Woman Ratio and General Fertility Rate

Another measure from a census or household survey is the child-woman ratio (CWR), the number of children under 5 divided by the number of women of childbearing age, multiplied by 1000. After division by five (because the numerator represents 5 years of births, rather than 1 year), the CWR can be interpreted as an estimate of the general fertility rate $2\frac{1}{2}$ years earlier, but with a downward bias because it omits children who died prior to the census or survey. It also slightly understates the number of women of childbearing age at the reference date because it omits women who died. If it pertains to a geographic subdivision of a country, it is affected by migration of mothers between the reference date and the survey date. The CWR is an indirect or substitute measure of fertility; variations in the CWR will correspond closely to variations in the direct measures of fertility.

The general fertility rate (GFR) is the number of births (in an interval of time) to women aged 15 to 49 (sometimes 15 to 44), divided by the total number of women aged 15 to 49 (or 15 to 44). Using woman-years, rather than numbers of women, a survey estimate of the GFR is simply the sum of the numerators of the age-specific rates, divided by the sum of the denominators of the age-specific rates:

$$GFR(t) = \frac{\sum_a b(a,t)}{\sum_a e(a,t)} \tag{16.8}$$

Using the 1998 NDHS of the Philippines, we estimate the GFR for 1990–1994 to be (710 + 2186 + 2370 + 1754 + 882 + 315 + 5)/(11,145.50 + 10,753.75 + 10,204.92 + 9040.25 + 7184.83 + 4789.58 + 364.58) = 8222/43,803.41 = 0.1877. The estimate for 1995–1999 would be (463 + 1342 + 1495 + 1098 + 676 + 211 + 24)/(8717.04 + 7091.21 + 6926.50 + 6589.42 + 5646.67 + 4469.67 + 2626.13) = 5309/42,066.64 = 0.1262. The numerators and denominators come from the last two columns of Tables 16.4 and 16.5, respectively.

Measures of Cohort Cumulative and Completed Fertility

Cumulative fertility refers to the number of children, or average number of children, born prior to some age or marital duration. Total or completed fertility refers to cumulative fertility up to the final age or marital duration in which any fertility occurs. This section will describe "true" cohort measures of cumulative fertility, "synthetic" cohort measures of cumulative fertility, and various linkages between these measures and the specific rates.

True Cohort Cumulative and Completed Fertility

Perhaps the simplest measure of cumulative fertility is the mean number of children ever born, or mean CEB. A question on CEB is included in every fertility or contraceptive prevalence survey, in many surveys that are conducted for entirely different purposes, and in many censuses.

The mean CEB for women within an age interval can be interpreted as the true cumulative fertility of a birth cohort. For example, the 1998 NDHS of the Philippines was conducted almost entirely during the months of March and April 1998, and the mean completed fertility of women aged 30 to 34 (at the date of interview) was found to be 2.69 children (Philippines NDHS Report, 1999, Table 3.6; weighted estimate). These women were born between the beginning of April 1963 and the end of April 1968. Making a coarse assumption of a uniform age distribution within the age interval 30 to 34, these women had an average exact age of 32.5 years at the date of interview. Thus, the figure of 2.69 is interpreted as the mean number of children born prior to age 32.5 by women who were themselves born from April 1963 through April 1968 and who survived to the date of interview. The cohorts represented by successive age intervals will slightly overlap because the field work is spread over an interval of time.

The women aged 35 to 39 in the same survey had a mean CEB of 3.47 children. It is not necessarily the case that 3.47 − 2.69 = 0.78 is the average number of children born between ages 32.5 and 37.5 for any real cohort, because the means 2.69 and 3.47 refer to different (even if slightly overlapping) birth cohorts. If fertility is increasing from one cohort to the next, as happened during the U.S. "baby boom," an older cohort may have lower cumulative fertility than a younger cohort, even at a later age.

To estimate the "current" (i.e., at the time of the survey) CEB for women at exact age 35, a researcher would typically average the means for ages 30 to 34 and 35 to 39, obtaining (2.69 + 3.47)/2 = 3.08. An alternative might be to calculate the mean for women in an age interval centered on exact age 35 (e.g., ages 32 years and 7 months through 37 years and 6 months of age) at the date of interview. For cohort comparisons, a more direct approach is possible with survey data. For the cohort born during 1960–1964, for example, exact age 35 was reached before the 1998 survey described earlier; so the birth histories could be used to calculate each woman's CEB at the beginning of the month when she had her 35th birthday. The average of these CEBs would be the average CEB at exact age 35 for this birth cohort. Birth cohorts above exact age 35 (at the survey date) could be compared in terms of the mean number of children they had had by exact age 35.

The CEB can be interpreted as *completed* fertility for women aged 45 to 49 or greater, but may actually underestimate completed fertility because of underreporting of the fertility of older cohorts. Women may omit children who were born long ago and died while young. Moreover, women who have had numerous children have a higher risk of dying from maternal or related causes. After the first birth, which is the most hazardous, the risk of maternal mortality is roughly proportional to the number of children. Women who died from such causes will be omitted from a census or survey, with the result that the mean CEB is biased downward. Thus, it is common in a less developed country for the mean CEB to reach a maximum for women aged about 45 to 49 and to decline steadily for older women, contrary to historical information about fertility trends.

An overall mean CEB calculated for women 15 to 49 will be sensitive to the age distribution within that age range. Many observed age distributions, particularly in the less developed countries, have more women in their twenties than in their thirties, and more in their thirties than in their forties. A mean will thus be weighted toward younger women, particularly if the population was growing rapidly when these women were born. A mean CEB calculated for a very broad age interval, or an overall mean CEB, is largely descriptive and has serious limitations for comparisons across groups or time periods. Direct standardization on some standard age distribution will slightly improve the usefulness of the CEB for making comparisons. In the absence of an obvious standard, a uniform age distribution can be used; in this case the standardized mean CEB will simply be the unweighted average of the mean CEBs in all the age intervals.

Synthetic Cumulative and Total Fertility

As a generalization, synthetic measures are constructed by interpreting period data as if they referred to a cohort. The best-known example is in the context of mortality (see, for example, Chapter 13), in which period data on the mortality of persons aged 0, 1–4, 5–9, . . . , 85+ are used to prepare an abridged life table. The survivorship column of the life table is interpreted as a description of a hypothetical or synthetic birth cohort as it passes from birth to exact ages 1, 5, 10, . . . , 85, even though the data for these age intervals actually come from different birth cohorts. The survivorship column of the life table provides a synthetic answer to a hypothetical "what if" question, namely "What is the chance of surviving to each exact age *a if* a cohort of women experiences throughout their lives the mortality observed in a recent interval of time?"

The concept of a synthetic cohort is easily extended to the measurement of fertility. (See, for example, Chapter 15.) Corresponding to the three period-specific fertility rates described earlier, for 5-year intervals of time and either age or marital duration, there are three cumulative totals for time period t, given as follows:

- $CFR(x,t)$ is five times the sum of the age-period-specific rates up to exact age x.
- $CMFR(x,t)$ is five times the sum of the age-period-specific marital rates up to exact age x.
- $CMDFR(x,t)$ is five times the sum of the marriage-duration-period-specific rates up to exact duration x

Thus,

$$CFR(x,t) = 5\sum_{a<x} f(a,t) \qquad (16.9)$$

This cumulative fertility rate at time t can be interpreted as the number of children that a woman would be expected to have (i.e., would have on average) if she experienced the time t rate for ages 15 to 19 for 5 years, the time t rate for ages 20 to 24 for 5 years, and so on, up to the age interval that extended from exact age x-5 to exact age x. In short, it is the expected number of births prior to age x, based on the fertility observed for different age intervals during time interval t. It is implicit that the woman survives to age x; possible mortality is ignored.

Similarly,

$$CMFR(x,t) = 5\sum_{a<x} mf(a,t) \qquad (16.10)$$

is the cumulative *marital* fertility up to age x, with the additional assumption that the woman is married from the earliest age in the summation, usually age 15. This cumulative rate can be very high because age-specific marital fertility rates are higher than age-specific fertility rates, especially in the younger ages where fewer women are married. It is preferable to apply a synthetic cohort interpretation to the *duration*-period-specific marital rates, in which case the cohorts are indexed by marital duration, rather than age. Thus, the cumulative marital duration fertility rate,

$$CMDFR(x,t) = 5\sum_{d<x} mf(d,t) \qquad (16.11)$$

gives the expected (or average) number of births in the first x years of marriage, without any reference, implicit or explicit, to age at marriage.

As described earlier, retrospective rates produced by a survey will be right-censored for earlier time periods, so that the full range of ages and marital durations will usually be available only for the most recent 5-year time period. For that time period, at least, the age-specific rates can be calculated out to ages 45 to 49 and the duration-specific rates out to duration 30 to 34. The cumulative fertility rate out to exact age 50 is the well-known total fertility rate, or *TFR*—that is,

$$TFR(t) = 5\sum_{a<50} f(a,t) \tag{16.12}$$

and the cumulative marital duration fertility rate out to exact duration 35 is known as the total marital duration fertility rate or *TMDFR*—that is,

$$TMDFR(t) = 5\sum_{d<35} mf(d,t) \tag{16.13}$$

There is no need to duplicate here the discussion of these rates and their interpretation given in Chapter 15. Also see Chapter 17 for modifications to limit the births to daughters and to take account of survivorship, leading to the gross reproduction rate and the net reproduction rate. Our purpose here is simply to indicate how the cumulative fertility rates, total fertility rates, and reproduction rates can be built up from specific rates derived from survey data.

Adding up the rates in the final columns of Tables 16.6 and 16.18, respectively, and multiplying by 5, we obtain an unweighted TFR for the Philippines in 1995–1999 of (0.0531 + 0.1892 + 0.2158 + 0.1666 + 0.1197 + 0.0472 + 0.0091) × 5 = 4.00. This number can be interpreted as the mean number of children that a woman would eventually have if she survived to the end of the childbearing ages and experienced the age-specific fertility rates observed during 1995–1999. The TMDFR for 1995–1999 is (0.3764 + 0.2204 + 0.1536 + 0.1051 + 0.0671 + 0.0299 + 0.0026) × 5 = 4.78. This number can be interpreted as the mean number of children that a woman would eventually have *if she ever married*, survived to the end of the childbearing ages, and experienced the observed *duration*-specific *marital* fertility rates.

The total marital fertility rate (TMFR) for 1995–1999, calculated from the final column of Table 16.12, would be (0.4475 + 0.3736 + 0.2706 + 0.1842 + 0.1284 + 0.0498 + 0.0096) × 5 = 7.32. This number can be interpreted as the mean number of children that a woman would eventually have if she married at age 15, survived to the end of the childbearing period, and experienced the observed age-specific marital fertility rates. The TMFR is required for the Bongaarts decomposition procedure (see Chapter 22). It should be interpreted cautiously because of its sensitivity to the high fertility of the early age intervals, even if very few women of those ages are actually married.

It is also possible to cumulate order-specific rates to obtain a total fertility rate TFR_j for each birth order *j*. For example, if the age-order-specific rates for order 1 are added across age (and multiplied by five if the age intervals are 5 years wide), we obtain TFR_1, which can be interpreted as the proportion of women who will ever have a first birth in a synthetic cohort subject to the observed period rates. It is possible for such a sum to exceed 1.0 if, say, the real cohorts tended to time their first births to occur during the observed period. For a real cohort followed over time, it would of course be impossible for the proportion to exceed one, so the interpretation must be modified if this happens to the synthetic measure. Similarly for higher birth orders. As noted earlier, if the age-order-specific rates are added across birth orders, we get the age-specific rates. Therefore, the sum of the order-specific total fertility rates, across birth orders *j*, will be the overall *TFR*—that is,

$$TFR_j(t) = 5\sum_{x<50} f_j(x,t) \tag{16.14}$$

and

$$TFR(t) = \sum_j TFR_j(t) \tag{16.15}$$

For 1995–1999, the order-specific rates for the illustrative data set are obtained from the last column of Tables 16.22 and 16.23 by adding the order-specific rates and multiplying the sums by five. They are as follows, for birth orders 1 to 4: $TFR_1 = 0.88$, $TFR_2 = 0.61$, $TFR_3 = 0.62$, and $TFR_4 = 0.47$. (Tables 16.22 and 16.23 give the order-specific rates for orders 1 and 2 only.) Calculated as a residual, $TFR_{5+} = 4.00 - 0.88 - 0.61 - 0.62 - 0.47 = 1.42$. For a synthetic cohort interpretation, about 88% of women would eventually have a first birth, about 61% would eventually have a second birth, about 62% would eventually have a third birth, and about 47% would eventually have a fourth birth. This kind of interpretation could be extended to individual birth orders five, six, and so forth, but not to an aggregation such as five or more.

Parity-Progression Ratios

Chapter 15 defined parity-progression ratios and showed how they can be calculated with vital statistics data. The birth histories in a fertility survey are a much more direct source of information about parity progression. The progression from parity *j* to parity *j + 1* is the closure of a birth interval, and birth histories contain information about both the beginnings and ends of birth intervals. The probability of making such a transition, given that parity *j* was reached (i.e., the parity-progression ratio) will be labeled PPR_j.

There are some important distinctions between order-specific fertility rates and parity-progression ratios, in terms of data requirements and interpretation, despite a superficial similarity in their names. Parity-progression ratios are indexed by the order of the birth that *begins* a birth interval (with a woman beginning at zero), whereas order-specific rates are indexed by the order of the birth that *closes* a birth interval. As another distinction, order-specific rates are typically calculated for specific ages (or marital durations) and periods; parity-progression ratios are typically calculated for cohorts or periods, but not for specific ages (or durations). Most important, parity-progression ratios are actually (estimated) conditional probabilities, rather than (central) rates, limited to the subpopulation at risk of making each successive transition to a higher parity.

True-Parity Progression Ratios

The following discussion draws on Hinde (1998, Chapter 9). If we follow a real cohort of women (that is, look retrospectively at the completed birth histories of the survivors of a real cohort), the probability that a woman who reached order j would go on to parity $j + 1$ could be readily estimated by dividing the number of women who ever reached parity $j + 1$ by the number of women who ever reached parity j. If n_k is the number of women who eventually had exactly k births, then

$$PPR_j = \left(\sum_{k>j} n_k\right) \Big/ \left(\sum_{k\geq j} n_k\right) \qquad (16.16)$$

Such an estimate can be seriously deficient if the cohort has not yet reached the end of the childbearing ages, because of two possible sources of bias. The first problem may be described as right-censoring. Some of the women who have reached parity j will eventually go on to parity $j + 1$, but they have not been observed long enough for this to be witnessed. Right-censoring always produces an *under*estimate of the cohort's eventual parity progression ratio. The second problem is left-censoring: some women have not yet even reached parity j, so the estimate will be biased toward women who reached parity j earlier, rather than later, in the life course. Women who reach a given parity early will tend to have larger completed families, so left-censoring tends to produce an *over*estimate of the cohort's eventual parity progression ratio.

Instead of regarding the parity-progression ratio as a characteristic of a birth cohort, with the attendant difficulty of making estimates before the cohort has completed its childbearing, we can shift to a period-specific definition from the birth-cohort definition, and try to estimate the probability of making a transition from parity j to parity $j + 1$, for all birth cohorts or age groups pooled, within an interval of time.

Suppose, for example, that we used the 1998 survey to estimate the progression from parity one to parity two. Pooling all age groups, we could identify women who had a first birth in 1990, say, and determine the proportion of them who had a second birth in 1990 or later. (It is possible to have two births in the same year, and we include the possibility of twins or other multiple births.) Then these women would have had nine calendar years (1990 to 1998; 1998 is only partially observed) in which to have a second birth. Since very few completed intervals are longer than this, the estimated PPR_1 could be interpreted as only a slight underestimate of the true probability that a woman who had a first birth in 1990 would eventually have a second birth.

Continuing to think of the transition from parity one to parity two, in order to keep the notation as simple as possible, let N_{1990} be the number of women who had a first birth in 1990, and *of those women*, let n_t be the number of women who had a second birth in t = 1990, . . . , 1997. Then

$$PPR_1 = \left(\sum_{k=0}^{8} n_{1990+k}\right) \Big/ N_{1990} \qquad (16.17)$$

The number of years following the reference year (in this case the reference year is 1990) is arbitrary, so long as it includes "virtually all" of the next-order births. To keep the right-censoring effect the same for a series of estimates, one could use 1980 through 1990, for example, as the reference years for the first birth and a 9-year interval (including the reference years) as the interval in which the second birth could have occurred.

Tables 16.28 and 16.29, from the 1998 NDHS of the Philippines, illustrate the necessary intermediate calculations. Table 16.28 shows that 337 women had a first birth in 1990. Table 16.29 gives the number of women who had a

TABLE 16.28 Births by Order and Calendar Year, 1990 to 1998: Philippines, 1998

Order	1990	1991	1992	1993	1994	1995	1996	1997	1998
1	337	366	402	331	389	385	406	416	85
2	327	323	341	339	294	301	346	339	78
3	268	270	268	292	259	263	237	273	68
4	221	214	212	216	191	202	187	186	42
5	173	133	157	156	154	147	133	144	36
6	108	99	124	108	105	111	94	118	25
7	80	68	75	57	80	82	76	65	23
8	59	58	56	37	50	52	50	54	11

Source: Based on 1998 NDHS of the Philippines (Philippines National Statistics Office, Department of Health, and Macro International, 1999).

TABLE 16.29 Number of Women With a Birth of Order 1 in Row Year and a Birth of Order 2 in Column Year, 1990 to 1998: Philippines, 1998

Year of order 1 birth	Year of order 2 birth								
	1990	1991	1992	1993	1994	1995	1996	1997	1998
1990	0	79	126	55	20	9	10	8	1
1991	0	0	101	121	40	25	18	12	1
1992	0	0	2	102	127	56	20	20	3
1993	0	0	0	7	65	107	50	22	8
1994	0	0	0	0	2	87	134	51	6
1995	0	0	0	0	0	1	93	119	17
1996	0	0	0	0	0	0	4	90	35
1997	0	0	0	0	0	0	0	5	5
1998	0	0	0	0	0	0	0	0	1

Source: Based on 1998 NDHS of the Philippines (Philippines National Statistics Office, Department of Health, and Macro International, 1999).

first birth in 1990 (or other calendar years) and had a second birth in a later year. The number of women shown for a second birth in years 1990 through 1998 is 0 + 79 + 126 + 55 + 20 + 9 + 10 + 8 + 1 = 308. Therefore the estimate of PPR_1 for reference year 1990 is 308/337 = 0.91. (This is probably an underestimate, since for earlier years we observe some longer gaps between first and second births.) Tables analogous to Table 16.29, describing transitions from a second to a third birth, from a third to a fourth birth, and so on, are also possible but are not presented here.

There is progressive left-censoring (omission of women who never had a first birth) in the estimates just described. It increases as we push the starting year backward in time because a fertility survey omits women over age 49 at the time of the survey. The 1990 estimate given earlier, for example, omits women who were over age 49 in 1998 (i.e., over age 41 in 1990) so the denominator of PPR_1 for 1990 is limited to women who had their first birth by age 41. The synthetic measures discussed here will reduce that problem.

Continuing to think of the transition from parity one to parity two, and retaining the previous notation for reference year 1990, the "true" probability of progressing from parity one to parity two,

$$PPR_1 = \left(\sum_{k=0}^{8} n_{1990+k} \right) \Big/ N_{1990}$$

is algebraically equivalent to

$$PPR_1 = 1 - (1 - a_0)(1 - a_1) \ldots (1 - a_8) \qquad (16.18)$$

where

$a_0 = n_{1990}/N_{1990}$
$a_1 = n_{1991}/(N_{1990} - n_{1990})$
$a_2 = n_{1992}/(N_{1990} - n_{1990} - n_{1991})$ and so on, and in general

$$a_j = n_{1990+j} \Big/ \left(N_{1990} - \sum_{k=0}^{j} n_{1990+k} \right)$$

Here a_j is analogous to q_x in the construction of a life table. It is the estimated probability of changing state (parity, rather than survivorship status) in an interval of time or age.

In words, the probability of *not* going on to a second birth (within 9 years of 1990, inclusive) is the probability of not going on to a second birth in the same year, times the probability of not doing so a year later (given that the woman did not already go on), times the probability of not doing so a year after that (given that the woman did not already go on), and so on, until "virtually all" transitions have occurred. Repeating the calculation of PPR_1 for reference year 1990, just described, this procedure would require the following intermediate steps:

$a_0 = 0/337$, $1 - a_0 = 337/337$
$a_1 = 79/(337 - 0) = 79/337$, $1 - a_1 = 258/337$
$a_2 = 126/(337 - 79) = 126/258$, $1 - a_2 = 132/258$
$a_3 = 55/(258 - 126) = 55/132$, $1 - a_3 = 77/132$
$a_4 = 20/(132 - 55) = 20/77$, $1 - a_4 = 57/77$
$a_5 = 9/(77 - 20) = 9/57$, $1 - a_5 = 48/57$
$a_6 = 10/(57 - 9) = 10/48$, $1 - a_6 = 38/48$
$a_7 = 8/(48 - 10) = 8/38$, $1 - a_7 = 30/38$
$a_8 = 1/(38 - 8) = 1/30$, $1 - a_8 = 29/30$

and

$$\begin{aligned} PPR_1 &= 1 - (337/337)(258/337)(132/258) \\ &\quad (77/132)(57/77)(48/57)(38/48)(30/38)(29/30) \\ &= 1 - (29/337) = 308/337 = 0.91 \end{aligned}$$

Note that this result is identical to the one obtained earlier.

Synthetic Parity-Progression Ratios

To construct a synthetic analog, we will index the measure by the year in which the *second* birth occurred and borrow the successive year-specific a_j measures from successive cohorts (indexed by the second birth) rather than from the same cohort (indexed by the first birth). As before, let N_t be the number of first births in year t. Expand the previous notation by replacing n_t with $n_{t1,t2}$, where $t1$ is the year when the first birth occurred and $t2$ is the year when the second birth occurred. Thus, the previous symbol n_{1991}, for example, would become $n_{1990,1991}$. The synthetic measure for 1997, the year of the *second* birth, would be

$$PPR_1^* = 1-(1-a_0^*)(1-a_1^*)\ldots(1-a_8^*) \qquad (16.19)$$

where again we cover a span of 9 years, inclusive, and

$a_0^* = n_{1997,1997}/N_{1997}$,
$a_1^* = n_{1996,1997}/(N_{1996} - n_{1996,1996})$
$a_2^* = n_{1995,1997}/(N_{1995} - n_{1995,1995} - n_{1995,1996})$ and so on, and in general

$$a_j^* = n_{1997-j,1997}\bigg/\left(N_{1997-j} - \sum_{k=0}^{j} n_{1997-j,1997-j+i}\right).$$

Tables 16.28 and 16.29 also include the necessary data from the Philippines' 1998 NDHS to estimate PPR_1^* for 1997 (indexed by the year in which the second birth occurred, rather than first). It requires these intermediate steps:

$a_0^* = 5/416 = 0.0120$
$a_1^* = 90/(406 - 4) = 0.2239$
$a_2^* = 119/(385 - 1 - 93) = 0.4089$
$a_3^* = 51/(389 - 2 - 87 - 134) = 0.3072$
$a_4^* = 22/(331 - 7 - 65 - 107 - 50) = 0.2157$
$a_5^* = 20/(402 - 2 - 102 - 127 - 56 - 20) = 0.2105$
$a_6^* = 12/(366 - 101 - 121 - 40 - 25 - 18) = 0.1967$
$a_7^* = 8/(337 - 79 - 126 - 55 - 20 - 9 - 10) = 0.2105$
$a_8^* = 1/(373 - 1 - 96 - 130 - 52 - 26 - 23 - 3 - 5)$
$= 0.0270$

$$PPR_1 = 1-(1-0.0120)(1-0.2239)(1-0.4089)(1-0.3072)$$
$$(1-0.2157)(1-0.2105)(1-0.1967)(1-0.2105)$$
$$= 0.88$$

This synthetic analog of the cohort estimate of PPR_1 is also biased downward somewhat because a few of the second births in 1997 were preceded by birth intervals longer than 8 years. Nevertheless, it is close to the true cohort estimate for first births in 1990, which was 0.91.

FINAL NOTE

The main goal of this chapter has been to describe in detail the manner in which a wide range of fertility measures can be calculated from survey microdata data. Nearly all of these measures were developed prior to the availability of fertility surveys and were originally defined in terms of vital statistics data for numerators and census data for denominators, as described in Chapter 15. The original definitions were appropriate for the available data sources, but—apart from limitations of sample size—retrospective surveys can be a superior source. When individual-level data are available, it is natural to see fertility as a stochastic characteristic of individuals (rather than as a deterministic characteristic of aggregates) to be expressed in terms of estimated expected values and estimated probabilities, conditional on a range of other characteristics. The essential ingredients are whether a birth (or a number of births) occurred in an interval, together with a measure of exposure to risk, such as the length of the interval or the amount of the interval spent in a given state (such as an age or marital status). Retrospective surveys fall short of a continuous population register, but they are much closer to the process than the traditional sources of data.

The individual-level components can be cumulated into numerators and denominators and manipulated to describe an aggregate, as in this chapter. They can also be used in multivariate statistical analyses that involve a wide range of risk factors, partitioning according to proximate determinants, and related variables and models. Poisson regression, logit regression, and hazard modeling are possible with the individual-level components, but these do not fall within the scope of this chapter.

References

Hinde, A. 1998. *Demographic Methods* (Chapters 8–11). London: Arnold.

International Union for the Scientific Study of Population (IUSSP). 1982. *Multilingual Domographic Dictionary*, English Setion. Second edition adapted by Etienne van de Walle. Liège, Belgium, Ordina Editions.

Philippines National Statistics Office (NSO) and Department of Health (DOH), and Macro International, Inc. (MI). 1999. *National Demographic and Health Survey 1998*. Manila: NSO and MI.

Shryock, H. S., J. S. Siegel, and E. G. Stockwell, 1976. *The Methods and Materials of Demography*. Condensed Edition. New York: Academic Press.

Suggested Readings

Bulatao, Rudolfo, A., and J. B. Casterling. (eds.). 2001. *Global Fertility Transition.* A Supplement to Vol. 27, 2001, *Population and Development Review.* New York: The Population Council.

Casterline, J. B., R. D. Lee, and K. A. Foote (Eds.). 1996. *Fertility in the United States: New Patterns, New Theories*. A Supplement to Vol. 22, 1996, *Population and Development Review.* New York: Population Council.

Cleland, J., and C. Scott (Eds.). 1987. *The World Fertility Survey: An Assessment of Its Contribution*. London: Oxford University Press.

Kiser, C. V., W. H. Grabill, and A. A. Campbell. 1968. *Trends and Variations in Fertility in the United States*. Cambridge, MA: Harvard University Press. Esp. pp. 297–310.

Whelpton, P. K., A. A. Campbell, and J. E. Patterson. 1966. *Fertility and Family Planning in the United States*. Princeton, NJ: Princeton University Press. Esp. pp. 1–31, 416–419.