CHAPTER THREE

RESEARCH METHODOLOGY

3.1. Introduction

This chapter deals with the methodological aspects of this study. It presents the source of data, the sample design, the instrumentation, the variables of interest of the study, and the data analysis technique.

3.2. Source of Data

This study is a secondary data analysis of the 2000 Rwandan Demographic and Health Survey (RDHS) of married people aged 15 to 49 years. This is the second round in a series of national-level Demographic and Health Surveys conducted in Rwanda under the worldwide Demographic and Health Surveys Program. The 2000 RDHS was conducted by the Rwandan Health Ministry in collaboration with the National Office of Population (ONAPO). ORC Macro provided technical support for the survey through the MEASURE DHS+ programme.

3.3. Sample design

The sample for the 2000 RDHS covered the population residing in private households in the country. A representative probability sample of about 9696 households was selected nationwide. The list of enumeration areas (EAs) from the 1997 Rwandan Finance Ministry’s Survey was used as a frame for the sample. All women aged 15-49 and all men aged 15-59 that were either usual residents of the households in the RDHS sample or visitors present in the household the night before the survey were eligible to be interviewed in the survey. In the households interviewed in the survey, a total of 10622 of eligible women aged 15-49 were identified. Interviews were completed with 10421 of these women, yielding a response rate of 98.1 percent. In the same households, a total of 2.857 eligible men aged 15-59 were identified and
interviews were completed with 2.717 of these men, yielding a male response rate of 95.1 percent.

3.4. Instrumentation

Three questionnaires were used for the 2000 RDHS: the Household Questionnaire, the Women’s Questionnaire, and the Men’s Questionnaire. The contents of these questionnaires were based on the model questionnaires developed by the MEASURE DHS+ programme and were designed to provide information needed by health and family planning programme managers and policymakers. The questionnaires were adapted to the Rwandan situation and were translated from English into Kinyarwanda, the local language. The Household Questionnaire was used to list all the usual members and visitors in the selected households. Information was collected on the characteristics of each person listed, including the age, sex, education, and relationship to the head of household. The main purpose of the Household Questionnaire was to identify eligible women and men for the individual interview. The Household Questionnaire collected information on characteristics of the household’s dwelling unit, such as the source of drinking water, type of toilet facilities, flooring materials, ownership of various consumer goods, and ownership and use of mosquito nets. It was also used to record height and weight measurements of women aged 15-49 and children under the age of 5, and to record the respondents’ consent to the haemoglobin and HIV testing.

The Women’s Questionnaire was used to collect information from all women aged 15-49. These women were asked questions on the following topics: respondent’s background characteristics, such as education, residential history, media exposure, knowledge and use of family planning methods, fertility preferences, antenatal and
delivery care, breastfeeding and infant and child feeding practices, vaccinations and childhood illnesses, childhood mortality, marriage and sexual activity, woman’s work and husband’s background characteristics, and awareness and behaviour regarding AIDS and other STIs.

The Men’s Questionnaire was administered to all men aged 15-59 in every household in the RDHS sample. The Men’s Questionnaire collected much of the same information found in the Women’s Questionnaire, but was shorter because it did not contain a reproductive history or questions on maternal and child health and nutrition.

3.5. Variables of interest

3.5.1. Dependent variable

There is only one dependent variable in this study, which is fertility. A number of measures of fertility are commonly used in population studies, including the crude birth rate (CBR), the general fertility rate (GFR), the age-specific fertility rate (ASFR), the total fertility rate (TFR), the gross reproduction rate (GRR) and the net reproduction rate (NRR). In this study, fertility is measured by the TFR and number of children ever born (CEB).

3.5.2. Independent variables

These consist of proximate and socio-economic variables that affect fertility behaviour in Rwanda. These include respectively: age at first marriage, contraception, breastfeeding; women’s education, husband’s education, women’s employment, infant survival, residence and religion. Among socio-economic variables, women’s education is the major independent variable in this analysis.
An examination of all these factors that affect fertility behaviour will therefore, enable an isolation of the net effect of women’s education and bring greater understanding to the dynamics of childbearing in Rwanda. The role of husband’s education is also examined because it could indicate the breaking of Rwandan men from pronatalist attitudes and other patriarchal traditions.

In fertility studies, the common approach in measuring the independent variables is by considering questions asked during the survey regarding the age at first marriage, the modern method of contraception used, the level of educational attainment or the highest level of education, the type of residence, etc. The table below presents how these proximate and socioeconomic variables can be measured.

**Table 1: Measurement of variables of interest in the study**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proximate variables</strong></td>
<td></td>
</tr>
<tr>
<td>Age at first marriage</td>
<td>Defined as such in variable list in the dataset( mean )</td>
</tr>
<tr>
<td>Contraceptive use</td>
<td>Ever used any modern method(yes, no)</td>
</tr>
<tr>
<td>Breastfeeding</td>
<td>Duration of breastfeeding(mean, median)</td>
</tr>
<tr>
<td><strong>Socioeconomic variables</strong></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>Highest educational level( no education, primary, secondary, higher)</td>
</tr>
<tr>
<td>Type of place of residence</td>
<td>Urban, rural</td>
</tr>
<tr>
<td>Employment status</td>
<td>Paid work, not paid work</td>
</tr>
<tr>
<td>Child survival</td>
<td>Child is alive( yes, no)</td>
</tr>
</tbody>
</table>

Source: drawn from DHS data information
3.6. Data analysis

In this study, three levels of analysis are considered. At one level, the analysis is limited to examine the respondent’s profile using univariate analysis. At the second level, a bivariate analysis enabled me to examine the relationship between the determinants of fertility in this study. The third level, concerned a multiple regression using path analysis to examine the contribution of women’s education (and other socioeconomic factors to test the net effect of women’s education) on fertility through the proximate determinants as shown in figure 2. In addition, estimations of fertility and the index of proximate determinants have been done. The choice of this statistical test is explained by the fact that unlike multiple regression in which each predictor variable has only a direct effect on the response variable; in path analysis predictor variables affect the response variable directly or indirectly through one or more intervening variables. Path analysis allows the simultaneous modelling of several related regression relationships where a variable can be a dependent variable in one relationship and the independent variable in another. The causal ordering as indicated in the figure 2 is derived from theory as it has been highlighted in the literature (Retherford & Choe, 1993).

Furthermore, regression analysis has been used to investigate the path model presented in figure 2. β coefficients from this analysis enabled me to compute the values that determine the direct and indirect effects of variables on fertility in the model. In this analysis, women aged 35-49 are examined because I assume that they have almost completed their childbearing.
In this model, SES is socioeconomic variable; A is age at first marriage; C is contraceptive use; B is breastfeeding, and F is total children ever born; b, h, d, j, f, l, are arrows showing the direction of the effect. The above figure represents a path diagram where arrows between variables indicate the direction of causation, and the numbers written on the arrows represent effects in the form of regression coefficients called path coefficients. As there are no arrows going towards socioeconomic variables (SES), that means they are not determined by other variables in the model, therefore, they are exogenous variables. Contrarily, F is an endogenous variable because all arrows come forward to it in the model. A, C and B are both endogenous and exogenous because there are arrows coming toward them and other arrows coming from them towards F. According to the above path diagram, regression equations for each endogenous variable can be written as follows:

<table>
<thead>
<tr>
<th>Model (a)</th>
<th>Model (b)</th>
<th>Model (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) A = a + b SES</td>
<td>F = g + h A</td>
<td>F = m + n SES + pA</td>
</tr>
<tr>
<td>2) C = c + d SES</td>
<td>F = i + j C</td>
<td>F = q + r SES + s C</td>
</tr>
<tr>
<td>3) B = e + f SES</td>
<td>F = k + l B</td>
<td>F = t + u SES + v B</td>
</tr>
</tbody>
</table>

1 We do not make use of model (b) because it concerns the part which has been highlighted in the previous section (proximate and fertility).
In order to find out the direct and indirect effects of socioeconomic variables, the equations in the model (a) can be substituted into equations in model (c) as follows:

4) \( F = m + n \text{ SES} + p (a + b \text{ SES}) = (m + pa) + (n + pb) \text{ SES} \):

5) \( F = q + r \text{ SES} + s(c + d \text{ SES}) = (q + sc) + (r + sd) \text{ SES} \)

6) \( F = t + u \text{ SES} + v(e + f \text{ SES}) = (t + ve) + (u + vf) \text{ SES} \)

The equations (4), (5), and (6) are **reduced form of the model** where the coefficients \( n, r, \) and \( u \) are **direct effects** of socioeconomic variables (SES) on fertility (F), and the product, \( pb, sd, vf \), are the **indirect effects** of socioeconomic variables (SES) on fertility (F). The **total effects** of socioeconomic variables on fertility equal the sum of the direct effects and the indirect effects (\( n + pb, r + sd, u + vf \)).

To calculate the effects of socioeconomic variables on fertility, the table below presents the numerical values of path coefficients generated from regression analysis. These coefficients are standardized because in this analysis, it is necessary to compare the effects of predictor variables on fertility. Furthermore, the coefficients of determination (R square) are also given in different models in order to assess the extent to which socioeconomic variables influence fertility through proximate determinants. As the substitution of equations from model (a) into model (c) are based on proximate determinants, equations (4), (5), and (6) represent three models corresponding to the three proximate variables namely age at first marriage, contraceptive use, and breastfeeding. The table below presents the model (A) including total children ever born, socioeconomic variables, and age at first marriage. Given that socioeconomic variables have been presented as a whole in the above models, it is necessary to specify them in the model in order to identify the effect of
each other on fertility. Therefore, specifying socioeconomic variables in models (a) and (c), the equations are as follows:

Model (A)

(a) \[ A_1 = a_1 + b_1 W_1; \quad A_2 = a_2 + b_2 W_2; \quad A_3 = a_3 + b_3 H; \quad A_4 = a_4 + b_4 R; \]
\[ \quad A_5 = a_5 + b_5 M \]

Where: \( a_1 \ldots a_5 \) are constants in the model (a);
\( b_1 \ldots b_5 \) are coefficients of socioeconomic variables in the model (a);
\( W_1, W_2, H, R, M \) are respectively women’s education, women’s employment;
Husband’s education, residence and infant mortality;
\( A_1 \ldots A_5 \) represent age at first marriage as a dependent variable in the model (a).

(c) \[ F_1 = (m_1 + p_1 a_1) + (n_1 + p_1 b_1) W_1; \quad F_2 = (m_2 + p_2 a_2) + (n_2 + p_2 b_2) W_2 \]
\[ F_3 = (m_3 + p_3 a_3) + (n_3 + p_3 b_3) H; \quad F_4 = (m_4 + p_4 a_4) + (n_4 + p_4 b_4) R \]
\[ F_5 = (m_5 + p_5 a_5) + (n_5 + p_5 b_5) M \]

Where: \( (m_1 + p_1 a_1) \ldots (m_5 + p_5 b_5) \) are constants in the model (c);
\( n_1 \ldots n_5 \) are numerical values of the direct effect of socioeconomic variables on fertility in the model (c);
\( p_1 b_1 \ldots p_5 b_5 \) are numerical values of the indirect effect of socioeconomic variables on fertility in the model (c);
\( (n_1 + p_1 b_1) \ldots (n_5 + p_5 b_5) \) are numerical values of the total effect of socioeconomic variables on fertility in the model (c). The same logic holds for the models B and C as shown in the chapter 6.

In this analysis, type of earnings for work (women’s employment) and type of place of residence have been used as dummy variables. The values of the two variables are related respectively to paid work (recoded 1) and urban residence (recoded 1)
compared to not paid work (recoded 0) and rural residence (recoded 0). The following chapters present the findings and the discussion of this study.