No. 9

Strategies for the Analysis of WFS Data

BASIC DOCUMENTATION

JANUARY 1977: WFS/TECH. 449

INTERNATIONAL STATISTICAL INSTITUTE
Permanent Office • Director: E. Lunenberg
428 Prinses Beatrixlaan
Voorburg, The Hague
Netherlands

WORLD FERTILITY SURVEY
Project Director:
Sir Maurice Kendall, Sc. D., F.B.A.
35-37 Grosvenor Gardens
London SW1W OBS, U.K.

The World Fertility Survey is an international research programme whose purpose is to assess the current state of human fertility throughout the world. This is being done principally through promoting and supporting nationally representative, internationally comparable, and scientifically designed and conducted sample surveys of fertility behaviour in as many countries as possible.

The WFS is being undertaken, with the collaboration of the United Nations, by the International Statistical Institute in cooperation with the International Union for the Scientific Study of Population. Financial support is provided principally by the United Nations Fund for Population Activities and the United States Agency for International Development.

This publication is part of the WFS Publications Programme which includes the WFS Basic Documentation, Occasional Papers and auxiliary publications. For further information on the WFS, write to the Information Office, International Statistical Institute, 428 Prinses Beatrixlaan, The Hague – Voorburg, Netherlands.

Strategies for the Analysis of WFS Data

Prepared by:
WFS CENTRAL STAFF
International Statistical Institute
35-37 Grosvenor Gardens
London SW1W OBS, U.K.

Contents

1	**Introduction**	5
2	**Evaluation of Data**	6
3	**The Unit of Analysis**	8
4	**Variables for Fertility Analysis**	9
4.1	Dependent Variables	9
4.2	Intermediate Variables	10
4.3	Explanatory Variables	11
5	**A Framework for Fertility Analysis**	21
	An Analytic Framework (figure)	
6	**Fertility Analysis and Some Lines of Inquiry**	22
6.1	Descriptive Analysis and the Measurement of Fertility	22
6.2	The Analysis of Fertility Differentials	23
6.3	Model Building in Fertility Analysis	24
7	**Some Notes on Methodology**	26
References		26
Table 1: Categories of Dependent Variables for Fertility Analysis		12
Table 2: Categories of Intermediate Variables for Fertility Analysis		14
Table 3: Categories of Explanatory Variables for Fertility Analysis		17

1. Introduction

The Guidelines for Country Report No. 1 set out the tabulations which can be derived from the WFS Core Questionnaire and the Fertility Regulation Module of the World Fertility Survey. The present document is a sequel to those Guidelines and is intended to indicate the directions in which further analytical studies may be pursued.

More detailed discussions on substantive and methodological issues arising in the analysis of WFS data subsequent to Country Report No. 1 will be published in a series of Technical Bulletins, which will complement this document and provide more specific guidelines for analysis.

Whereas in the Guidelines for Report No. 1 it was possible to set out the recommended tabulations in considerable detail, this document must necessarily be of a less definite kind. Some of the topics considered below may not be of interest to a particular country; others may require study in greater depth. This document should therefore be regarded as a general strategy and some suggestions on topics for analytical study in the individual country.

Although Report No. 1 will consist mainly of a description of the survey and an evaluation of the techniques used in carrying it out, together with the tabulations, the preparation of that report will probably have suggested hypotheses for examination in the light of the WFS data. Those responsible in the country concerned will wish to consider all these matters in conjunction with the topics suggested below and to decide which of them they wish to work upon. The WFS staff will be glad to advise or to help in making this choice.

Apart from the choice of individual topics, there are certain general matters requiring examination before analyses are begun.

a) Following anything which may have been said in Report No. 1, an elaboration of issues relating to data evaluation, such as a discussion of sources of biases and errors in the data themselves.

b) An examination of the question whether any adjustments to the data are desirable or necessary, for example whether imputation of missing values, or whether corrections for bias in reporting birth intervals are required. If any data adjustment of such a kind is carried out, it is a basic rule that there should be a clear specification of what has been done and sufficient unadjusted data given to permit an independent alternative analysis to be performed.

c) A consideration of the question whether the data are too inadequate, by reason of quality or paucity, to bear refined analysis.

d) An examination of what techniques are going to be used in the analysis and of the computational effort which will be required.

e) A consideration of the time, cost and staff required in the proposed studies.

2. Evaluation of Data

As the WFS is aiming both to normalize procedures and to achieve higher standards of performance, data evaluation is of major importance. Thus the reporting of deficiencies as a guide to future work should be systematically undertaken (and, may form a separate document) and is in no way indicative of inadequate field work or machining. In this connection Occasional Paper No. 15 which deals with the problems encountered in the Fiji Fertility Survey[1] may be of interest.

A discussion of quality control, especially in the context of **nonsampling errors** will in most cases be published in Report No. 1. Further consideration of such matters as experience with spot-checks, field editing, office editing, coding, tape recording, and re-interviewing of subsamples where a post-enumeration survey is undertaken may be useful and could conveniently be incorporated in the separate report mentioned at the beginning of this section.

As soon as a clean data tape has been produced, it is essential to subject the data to tests designed to evaluate their overall quality, in particular the reliability of ages, dates and intervals in the maternity history. The nature of the tests may vary from one country to another. The WFS is working in collaboration with a number of institutions, in particular the London School of Hygiene and Tropical Medicine and the Office of Population Research at Princeton, to develop some useful tests and to publish recommendations at the earliest possible date. The result of these tests may determine the nature of the analyses to which the data may usefully be subjected. For example, in a country whose maternity history data appear to be badly defective, the analyses might be based only on children ever born; where the data are of high quality, a full analysis could also be conducted based on birth intervals. Intermediate between these extremes there may be cases where the defects are not serious enough to warrant rejection, but nevertheless require some modification in the analysis.

Again, comparisons can usefully be made on certain items between the Household Schedule and the Individual Questionnaire. The respondent giving the information for the first may not be the same as the respondent for the second; and the two may be interviewed at different times. Where discrepancies between the two sources are discovered some method of reconciliation is required, or at least a study of significant differences. Such a study depends on the extent of the discrepancy. If the differences are small, some simple method such as choosing one of the sources may suffice; in other cases, more sophisticated techniques may be appropriate.

It is also important to check the WFS data against external sources of information, for example, census figures or previous social surveys. Observed WFS patterns can be compared *inter se*, with what should be expected from the logic of the situation and from known empirical relationships found in accurately reported data.

It is obvious that WFS data must meet minimum standards of reliability before more detailed

1 M. A. Sahib *et al.*, *The Fiji Fertility Survey: A Critical Commentary on Administration and Methoddology*, WFS Occasional Paper Series (The Hague: International Statistical Institute, 1975)

analyses are based upon them. In some cases, adjustments may be made, but a point raised earlier in this document is worth re-emphasis. When adjustments are made to the primary data, the report should describe their nature and extent, and a record of the unadjusted data should be kept. This point will assume particular importance if different researchers work on the same material, especially in the matter of inter-country comparisons.

An extreme case arises when some data are missing. It is impossible to prescribe general rules as to when a record is so incomplete as to be not worth retaining. Specific decisions will depend on the kind of information which is missing and how relevant it is to the particular topic under study. In general, however, a distinction can usefully be made between straightforward tabulation and more sophisticated analytical techniques. In the former case the WFS is, in general, opposed to imputing missing values except in a few cases such as month of birth where only the year is given. Missing values in tabular work can usually be taken into account by columns headed "Not Available", or some similar term; or by a footnote recording the proportion of missing values in the total sample. For more sophisticated routines involving numerical operations on the data, missing values are a serious nuisance and some systematic imputation may be necessary. Should the need arise this subject might be taken up in the series of Technical Bulletins.

As concerns **sampling error** (namely uncertainty due to the fact that only a sub-set of the whole population has been subject to inquiry) the WFS is preparing a machine program for general use to enable standard errors, the effect of clustering, and the efficiency of the sample design to be calculated. An account of the theory involved and a number of illustrative examples are given in Occasional Paper No. 17.[2] Futher material will become available as surveys are completed.

It is recommended that sampling errors should be calculated of means and of proportions in important classes or sub-classes of the most important variables; although it does not follow that they should all be published. Any uncertainty in the figures should be dealt with in the text which considers the conclusions to be drawn from them.

Finally, it must be emphasized that evaluation is a continuing process and should be borne in mind at all stages of analysis.

2 L. Kish *et al*, *Sampling Errors in Fertility Surveys*, WFS Occasional Paper Series (The Hague: International Statistical Institute, 1976)

3. The Unit of Analysis

Before discussing the analysis of WFS data it is convenient to mention briefly the nature of the available data and to define the basic unit of statistical analysis.

Most of the WFS data pertain to characteristics of individuals, and therefore the basic **unit of observation** is the individual respondent. In a number of countries, however, data have also been collected on characteristics of households and communities. Thus, in some instances there may well be two or three types of units of observation and two or three different types of data available.

These different types of data may be integrated at the individual level, however, by assigning to each individual respondent the characteristics of the household and the community where the individual lives. This integrated approach has the advantage of making all data available at the level of the smallest possible unit, and thus provides flexibility in the selection of analytic levels and techniques.

These individual records are the input of the statistical analysis, as they provide a basis for the construction of tables and the computation of various kinds of statistics. Thus the basic **unit of analysis,** in the statistical sense of this term, is the individual respondent. The assessment of the fertility levels of several ethnic groups, for example, may involve the computation of mean parity by age or marital duration on the basis of individual records.

The substantive interpretation of the results, on the other hand, depends on the nature of the variables involved and the theoretical framework being used. A variable such as ethnicity, for example, classifies each individual respondent as a member of a group with well-defined characteristics, and the analytic interpretation of the results may emphasize how these characteristics affect the fertility of the group as a whole. In this case, it is the group rather than the individual which constitutes the unit of analysis in a substantive sense. To avoid confusion in terminology the group may be referred to as the **unit of reference.**

The approach adopted in this document is that while the unit of reference may vary from one analysis to another, the unit of analysis will usually be the individual respondent. Therefore, the variables for fertility analysis, which are discussed in the next section, have been defined in terms of observable characteristics of individuals.

In addition to the types of data mentioned so far, in some situations it may be possible to obtain household or community data by aggregating the information for the individuals interviewed in each household or community, and to use these aggregate data as the basic input for the statistical analysis. This possibility, however, is likely to be limited by the nature of the sampling frame, and may also create statistical problems because of the various ways in which aggregation affects estimates and their standard errors.

4. Variables for Fertility Analysis

The analyst of fertility is faced with a complex multivariate system in which variables are likely to be interdependent. For example, it is well known that current fertility is dependent on the parity already achieved, but at the same time parity is itself a conventional dependent variable.
Nevertheless, in order to put some structure into the analysis, it is convenient to divide the variables into three groups, following Davis and Blake (1956):

a) The *dependent* variable, fertility itself;
b) *Intermediate* or *proximate* variables which have a direct bearing on fertility, such as exposure to intercourse, use of contraception and outcome of gestation;
c) *Explanatory* variables which are believed to affect fertility but in a more round about way through their bearing on the intermediate factors, and which include social and economic factors and attitudes to family size and structure. The explanatory variables themselves may be interlinked. One of the most difficult parts of the analysis is to dissect, if possible, their relative influence through the intermediate variables on the dependent variables.

It is customary in the analysis of fertility to employ several individual attributes as **controls,** in the sense in which this term has currency in demography. Their designation as control variables arises from their close logical or even biological relationship with the dependent variable. The most frequently used controls, employed either singly or in combination, are age and duration of marriage. Other variables, such as parity, are also frequently used.

4.1 DEPENDENT VARIABLES

The generic variable **fertility** has been defined as the production of live births by individuals, couples or populations, (see for example Grebenik and Hill, 1974). This definition covers a complex process extending for the individual woman through the many years in her lifetime in which she may bear children, and encompassing, for whole populations, the formation and dissolution of unions in which women are exposed to the risk of child-bearing. The process cannot be defined in terms of a single *dependent* variable applicable to all instances of fertility analysis.
The generic concept fertility is, however, illuminated by employing as dependent variables in the analysis a number of measureable quantities, each related to a particular aspect of this complex process. A list of such variables is presented in the first column of Table 1, page 12. In the second column of this table, reference is made to the items in the WFS Core Question-

naire that serve as the basis in constructing them, and in the third column a brief commentary on their possible use in quantitative analysis is given.

Problems that arise in the quantitative analysis of fertility are discussed in Section 6 below.

A variable that is hard to classify as dependent or intermediate is **fecundability**, which is defined as the probability of a woman conceiving in a menstrual cycle. The UN Multilingual Demographic Dictionary (UN 1958) and the International Demographic Terminology (IUSSP: Grebenik and Hill, 1974) have restricted the concept to situations where no contraception has been practised. Some demographers, however, find it useful to distinguish two types of fecundability – **natural fecundability** (in the absence of contraceptive practice) and **fecundability** (irrespective of whether contraception is practised or not).

According to either definition, fecundability could be estimated from data on the timing of unions, pregnancies and births, used in conjunction with information on other factors such as the duration of post-partum amenorrhoea and incidence of foetal loss. The data on reproductive history, particularly where the "Factors other Than Contraception" module is used, could be exploited for the estimation of fecundability for different populations.

4.2 INTERMEDIATE VARIABLES

The *intermediate* variables used by the WFS are classified in Table 2, page 14, in terms of the widely-known Davis-Blake framework.

This framework comprises a list of eleven mutually exclusive and exhaustive **categories**, each of which contains one or more variables. All forces affecting fertility must operate through one or more of the intermediate variables.

The eleven categories fall in turn into three **classes**, depending on whether they relate to exposure to intercourse, to conception (assuming intercourse is occurring), or to gestation and parturition (assuming conception has occurred).

The Davis-Blake categories provided in Column 1 of Table 2 (in contrast to the corresponding Column of Table 1) constitute a fairly general taxonomy rather than a list of specific WFS Questionnaire items. Futhermore, as parts of a purely logical classification scheme, the different categories have a relative importance that remains unspecified and that varies from one setting to another.

Although in most country surveys the WFS is collecting at least some information that falls within each of the appropriate Davis-Blake categories, there is no attempt in any single survey to obtain exhaustive or even maximal data on intermediate variables.

In the past analysts have often been forced by data restrictions to by-pass the intermediate variables, limiting their analysis to the relation between fertility and observed *explanatory* variables; others have concentrated on the intermediate variables to the exclusion of the more causally remote explanatory variables. The WFS is collecting information on all three kinds of variables in the expectation that a thoroughly penetrative analysis will probably need to make use of them all.

There may be occasions when some of the intermediate variables will be considered as *dependent* and will be related to other data, such as some of the explanatory variables; if the relationship is close the latter may be used as substitutes for the former, recognizing that some links in the causal chain have been suppressed.

4.3 EXPLANATORY VARIABLES

For WFS purposes it is convenient to divide the *explanatory* variables into five categories, which are ordered roughly from causally proximate to causally more remote:

a) Knowledge of and attitudes towards contraception
b) Attitudes relating to family size, structure and formation
c) Socio-economic and cultural characteristics
d) Bio-social characteristics
e) The socio-economic structure and the environment

Table 3, page 17, sets out some of the explanatory variables which, it is expected, would be used in analytical studies of the WFS data, together with a reference to their source in the WFS Core Questionnaires and a brief commentary on their use. The list does not purport to be exhaustive, the emphasis being placed on illustrating the types of data available rather than providing a complete list of variables explanatory of fertility.

Most of the variables in the list are included in the WFS Core Questionnaire and can be treated either on the basis of responses to individual questions or by amalgamating the responses to a series of questions, as for example in the case of pattern of work. (For definitions of constructed variables see Guidelines for Report No. 1).

There remains for consideration a number of variables which may not appear in the Core but are introduced into particular questionnaires by way of modules, or otherwise. These include socio-economic variables such as family income, cultural variables such as religion; and contextual or community-level variables relating to areas rather than individuals. Here some material is available in the WFS publications (e.g., on community variables by Ronald Freedman and on economic variables by Deborah Freedman and Eva Mueller)[3] but implementation at the analytic stage has not been fully worked out. The WFS may commission studies to explore the methodology and to offer guidelines or references to existing work on the subject.

The amalgamation of groups of variables into single measures, scales or index numbers is still to some extent an unexplored subject; but there again WFS may commission studies of the necessary methodology.

There may be occasions when some of the explanatory variables will become dependent and will be related to fertility or some of the intermediate variables. For example, attitudes to family size may be studied as a function of achieved parity.

3 Ronald Freedman, *Community-Level Data in Fertility Surveys* and *Examples of Community-Level Questionnaires*, WFS Occasional Paper Series Nos. 8 and 9; Deborah Freedman and Eva Mueller, *Economic Data for Fertility Analysis* and *Economic Modules for Use in Fertility Surveys in Less Developed Countries*, WFS Occasional Paper Series, Nos. 11 and 12, (The Hague: International Statistical Institute, 1974)

TABLE 1

CATEGORIES OF DEPENDENT VARIABLES FOR FERTILITY ANALYSIS*

VARIABLE	SOURCE	NOTES
I. VARIABLES BASED ON NUMBER OF EVENTS		
1. Number of children ever born	Q.211 in maternity history	This is the classic *dependent* variable in fertility analysis. Since parity is affected by total time exposed, care must be exercised in the analysis to introduce proper controls for length of exposure, through age or preferably marital duration. If the analysis is restricted to women 45–49 (or 40–49) years old, this variable represents completed fertility; unfortunately, this is a small group and its experience will reflect past rather than current trends in fertility.
a. Total		
b. Surviving	Qs. 201–207, 210, 211, 214 215 in maternity history.	These two variables are not fertility variables in a strict sense, but they are used to estimate fertility by methods such as the "own children" technique.
c. Own children living with respondent.	For **all women,** also available Cols. 11–17 in Household Schedule.	
2. Number of live births in past 5 (or 3) years.	Q.212 in maternity history	This variable represents an attempt to focus the analysis on recent fertility. The reference period used should be preferably 5 and in no case less than 3 years, because both sampling and response errors become unacceptably large if shorter periods are used. Unfortunately, this entails some loss of currency in the results. Control for exposure should be introduced in the analysis, for example restricting consideration to women married at least 5 (or 3) years.
3. Number of live births in first (second) 5 years of marriage.	Q.212 maternity history, and Qs. 403, 409 in marriage history.	These variables may be used to study the timing of fertility throughout marital life. The reference period of 5 years is recommended as a compromise: a long reference period is desirable to obtain reliable data, but a short reference period is needed to obtain resolution in the analysis. Controls for exposure should be introduced, for example by restricting the analysis to women married at least 5 (or 10) years. If the data bear it, shorter periods may be used.

II. VARIABLES BASED ON INTERVALS BETWEEN EVENTS.		
4. Length of the open interval.	Q.212 in maternity history.	The interval since the last birth (or since marriage, if the woman has had no births) provides a sharp focus on current fertility. This variable has been used both in mathematical models and empirical studies of fertility. Consideration must be given in the analysis, as with all interval data, to the effects of censoring and truncation.
5. Length of the closed interval.	Q.212 in maternity history.	The interval between the last two births (or between marriage and the first birth, if the woman has had only one birth), provides another indicator of recent fertility. In the case of women with no births as of the time of the interview this variable is undefined.
6. Interval between first marriage and first birth.	Q.212 in maternity history and Qs. 403, 409 in marriage history.	This variable may be used to study fertility in early marital life and is of considerable practical interest. The interval is censored for women who have not had their first birth by the time of the interview and this fact must be considered in the analysis, using for example life-table techniques.
7. Intervals between successive births.	Q.212 in maternity history and Qs.403, 409 in marriage history.	Used in conjunction with interval from marriage to first birth, the interval between first and second birth provides information on the timing of fertility in early marital life. The variable is censored for women who have had only one birth and undefined for women who have had no births. Other variables of interest are the intervals between second and third birth, between third and fourth, and so on. These intervals may be analyzed in a similar fashion as long as the number of cases does not become too small.

* These variables are numbers of events or lengths of intervals. When used in the computation of measures of fertility to be employed in analysis, they must of course be related to a population at risk or other appropriate denominators.

TABLE 2

CATEGORIES OF INTERMEDIATE VARIABLES FOR FERTILITY ANALYSIS

VARIABLE	SOURCE	NOTES
I. FACTORS AFFECTING EXPOSURE TO INTERCOURSE		This first group of variables deals with marital exposure. The term *marriage* is used here in a broad sense to include all sexual unions. Although every society has a type of union where reproduction is expected and approved, there may be other types of unions where reproduction may occur, such unions may be frequent and must be considered in the analysis. In addition to the intermediate variables mentioned below, the WFS collects data on number of unions, type of dissolution of each union and (optionally) type of union. Although these might be considered explanatory variables, they are mentioned here because they all deal with the formation and dissolution of unions.
A. **Formation and Dissolution of Unions**		
1. Age at entry into sexual unions.	Q.403 or 409 in marriage history.	This is a very important variable in fertility analysis as it marks the beginning of marital exposure. On subtracting it from current age we obtain marital duration which is used frequently as a demographic control representing roughly time of exposure. It can also play an explanatory role, as in the analysis of interval from marriage to first birth. Finally, it will frequently be analysed as a dependent variable in its own right; in such a case careful consideration must be given to the truncated nature of data on age at marriage based on a sample of ever-married women.
2. Permanent celibacy	Household Schedule, Columns 5 and 9.	Although data on women who do not marry by age 49 are available from the Household Schedule, the primary study population for WFS purposes is often restricted to ever-married women and thus this variable will not be considered in many analyses.*
3. Amount of reproductive period spent after or between unions.	Qs. 403, 406, 409, 411, 412 in marriage history.	This variable may be determined from the dates of beginning and end of each marital exposure as registered in the marriage history. On subtracting this from marital duration, we obtain total length of marital exposure. Although this provides a more refined demographic control than marital duration, considerations of simplicity and reliability of data lead frequently to the use of marital duration.
B. **Exposure to intercourse within unions.**		This group of variables deals with exposure to intercourse within sexual unions. As a group, they are much more difficult to ascertain than varia-

		bles in the first group. Not surprisingly, therefore, a more thorough coverage of them has been relegated from the Core Questionnaire to the module on "factors other than contraception affecting fertility".
4. Voluntary abstinence	Q.311 in use of contraception, and the module Factors other than Contraception Affecting Fertility.	This variable includes abstinence as a means of contraception which is considered in the Core Questionnaire, as well as post-partum abstinence, "occasional" restrictions on intercourse, gestational abstinence, prohibition of coitus during menstruation and prolonged abstinence in the last stages of reproductive life. The most important of the latter variables is post-partum abstinence, which is included in the module on factors other than contraception affecting fertility.
5. Involuntary abstinence	Module on Factors other than Contraception Affecting Fertility.	This includes abstinence due to impotence, illness and unavoidable but temporary separations of the marriage partners; the latter are considered in the module on factors other than contraception affecting fertility, the former are not ascertained. No consistent overall differences in fertility have yet been attributed to involuntary abstinence in the literature, except in the case of involuntary abstinence resulting from migration, but this may well be due to a lack of data.
6. Coital frequency	Module on Factors other than Contraception Affecting Fertility.	This variable excludes periods of abstinence. Existing data on coital frequency do not support a consistent view, are ambiguous and of doubtful reliability; thus the value of this variable in the analysis of WFS data cannot be predicted.
II. FACTORS AFFECTING EXPOSURE TO CONCEPTION		This group of variables deals with exposure to the risk of conception given exposure to intercourse, thus dealing with the second necessary step in the process of reproduction. These variables are of considerable importance in the determination of fertility levels in many societies and have therefore been included in the Core Questionnaire.
7. Fecundity or infecundity as affected by involuntary causes.	Qs.504, 505, 509, 510, 511, 512, 525, 526 in fertility regulation section, see also the module Factors other than Contraception Affecting Fertility.	This variable includes female sterilization for purposes other than contraception and other fecundity impairments which are considered in the Core Questionnaire. We also include post-partum amenorrhoea and lactation as they affect fecundity, these variables being considered in the module on factors other than contraception. Lactation may be affected by the survival of the preceding child (see Table 3, Variable 18).

* This procedure differs where all women are eligible to be interviewed.

TABLE 2 (Continued)

VARIABLE	SOURCE	NOTES
8. Use or non-use of contraception.	Qs.304–314, see column 3 in section on contraceptive knowledge and use, and Qs.504, 505, 507, 508, 529, 530 in fertility regulation section.	A very important variable in fertility analysis. The data obtained in the WFS Core Questionnaire include ever-use of contraception as well as use of contraception at different stages of marital life, such as currently and in the open interval. Sterilization and abstention are not considered contraceptive methods in this scheme, but are included in variables 7, 9 and 4. Further information is obtained in the Fertility Regulation Module.
9. Fecundity or infecundity as affected by voluntary causes.	Qs.504, 505, 509, 510, 511, 512, 525, 526 in fertility regulation section.	This includes female sterilization for contraceptive purposes and male sterilization, which in the WFS Core Questionnaire is assumed to be always for contraceptive purposes.
III. FACTORS AFFECTING GESTATION AND SUCCESSFUL PARTURITION		This last group of variables deals with gestation and parturition given exposure to conception, the third necessary step in the reproductive process. Pregnancy wastage is probably an important factor in the determination of fertility levels and differentials, but it is rather difficult to assess reliably. Emphasis in the Core Questionnaire has been placed on the assessment of fertility; a thorough assessment of pregnancy wastage requires use of the Abortion Module.
10. Foetal mortality from involuntary causes	Qs.225–228 and Qs.229–232 in maternity history, and Abortion Module.	Although a list of other pregnancies is obtained in the Core Questionnaire, no attempt is made to distinguish voluntary from involuntary foetal mortality. Hence, unless the Abortion Module is used, this variable cannot be analysed separately from the next one.
11. Foetal mortality from voluntary causes.	Qs.225–228 and Qs.229–232 in maternity history, and Abortion Module.	The effect of this variable on fertility is really dependent upon whether or not abortion is practised as a means of fertility regulation. As a voluntary factor it is potentially important in the production of fertility differentials.

TABLE 3

CATEGORIES OF EXPLANATORY VARIABLES FOR FERTILITY ANALYSIS

VARIABLE	SOURCE	NOTES
I. KNOWLEDGE OF AND ATTITUDES TOWARDS CONTRACEPTION		The first and most obvious group of *explanatory* variables deals with knowledge of contraception, which affects fertility via intermediate variables 8 and 9, the use and non-use of contraception and sterilization for contraceptive purposes, and may be affected in turn by other, causally more remote, explanatory variables.
1. Knowledge of specific methods	Qs.301–307 in the section on contraceptive knowledge and use.	The WFS Core Questionnaire collects information on knowledge of several specific methods such as the pill, IUD and others. Knowledge of male and female sterilization is also included here.
2. Knowledge of efficient methods	Qs.301–307 in the section on contraceptive knowledge and use.	Of particular interest is the knowledge of any modern, efficient contraceptive, capable of a significant impact on fertility, and how this knowledge may depend on social and economic variables.
3. Attitudes towards contraception	Family Planning Module	Attitudes towards contraception affect use and hence fertility, and are in turn dependent on a number of social and economic variables (e.g., religion).
II. ATTITUDES RELATING TO FAMILY SIZE, STRUCTURE AND FORMATION		This second group of explanatory variables deals with the respondent's fertility intentions and ideals, a more elusive set of variables. Their effect on fertility may be mediated by several intermediate variables including 1, 4, 8 and 11, i.e., age at marriage, voluntary abstinence, use or non-use of contraception and induced abortion. One area for research is the extent to which these intentions are translated into consistent behaviour; another is the extent to which the intentions are affected by socio-economic and cultural variables.
4. Desire for a future birth	Qs.514, 517, 520 in the fertility regulation section	An expression of current intentions which may be studied as a function of achieved fertility and past intentions as well as an explanation of current behaviour.
5. Additional number of children desired	Qs.516, 519, 521 in the fertility regulation section	A more quantitative expression of current intentions. May be analyzed in the same fashion as the previous variable.

TABLE 3 (Continued)

VARIABLE	SOURCE	NOTES
6. Desire for last birth or current pregnancy	Fertility Regulation Module	The question of desire for more children is pursued in more detail in the module by going back to the last birth. This may be related to contraceptive behaviour prior to the last birth, data on which are also collected in the module.
7. Ideal number of children	Q.531 in fertility regulation section	This variable attempts to get at preferences for number of children independently of achieved fertility, and may reflect social norms as perceived by the members of society.
8. Preference for a boy or a girl	Qs.515, 518, 223 in fertility regulation section and maternity history	It has been hypothesized that the desire to have a boy or a girl may affect fertility. This variable provides a measure of gender preferences for women who want another child, and may be studied as a function of the sex of the children already had.
9. Perception of economic costs and benefits of children	Economic Module	Fertility intentions may be affected by the perceived advantages and disadvantages of children, including their economic impact in the family. The assessment of these costs and benefits is considered in a special module.
III. SOCIO-ECONOMIC AND CULTURAL CHARACTERISTICS		The third group of explanatory variables deals with social, economic and cultural characteristics of individuals. These are affected by the social and economic structure, and in turn affect fertility intentions and the knowledge of contraception, thus having an indirect effect on fertility via several intermediate variables; they can also have a more direct effect mediated by variables such as 1 or 8, age at marriage, and use or non-use of contraception.
10. Migration status	Q.104	Migrants may well have different values and behaviour patterns from the native born. Information on migration status might be supplemented by optional questions on birth-place and length of residence.
11. Childhood type of place of residence	Qs.105–106	People who grow up in the countryside, small towns or cities are exposed to different values, norms and socialization experiences, which contribute to mould their attitudes and affect their behaviour later in life.

12. Religion, Ethnicity, Language	Optional in the Core Questionnaire	Religious, ethnic and linguistic groups also differ in norms and values, and frequently vary in their reproductive behaviour. In countries where different religions, ethnic groups or linguistic groups are found these are important characteristics to consider.
13. Education	Qs.109–111, 113 in respondent's background	Existing data indicate the almost universal existence of fertility differentials by education; the exact form of the relationship varies however, and the mediating causal mechanisms are complex and worthy of detailed study.
14. Pattern of work	Qs.614–617, 610, 612, 602–604, 608–609 in woman's work history	The woman's participation in the labour force at different stages in her reproductive life is another important variable in fertility analysis, affecting her attitudes and behaviour.
15. Family income	Economic Module	The assessment of income is complex and has therefore been considered in a special module. This is a very useful complement of other indicators of socio-economic status.
16. Background of husband	Section 7 on current or last husband's background	We consider here a cluster of characteristics which includes husband's childhood type of place of residence, education and occupation. An area for research is the interaction of the husband's and wife's characteristics in the determination of the couple's fertility.
IV. BIO-SOCIAL CHARACTERISTICS		The fourth group of explanatory variables deals with bio-social factors which can affect fertility more directly through intermediate variables such as 5, 7 or 10 (abstinence, fecundity or infecundity, and foetal mortality) all from involuntary causes, and less directly through other explanatory variables.
17. Nutrition and Health	For the present, the Factors other than Contraception Module (including height and weight) and data from extraneous sources	The woman's nutritional level, and more generally her health status, can affect directly intermediate variables such as 5, 6 and 7, which include involuntary abstinence due to illness, coital frequency and fecundity as affected by post-partum amenorrhoea or lactation, and may be important in many societies.
18. Infant and childhood mortality	Qs.214, 215 in maternity history	It has frequently been hypothesized that high fertility may be due to high general and child mortality levels, and more specifically that parents will either try to replace the children who die or anticipate the toll of mortality by having more than desired. The extent to which this obtains and the study of the mediating mechanisms are important areas for

TABLE 3 (Continued)

VARIABLE	SOURCE	NOTES
		research. For populations having long durations of lactation, deaths of live born infants prior to two years of age are equivalent to weaning, in which case it may be useful to consider these as intermediate variables. In particular, neo-natal infant mortality is closely related to late foetal mortality (already classified as intermediate variable 10 in Table 2), as a component of perinatal mortality.
V. THE SOCIAL STRUCTURE AND THE ENVIRONMENT		The fifth and last group of explanatory variables deals with the community, the society and the environment. Some of these variables are more remote in the causal chain, exerting most of their effect on the intermediate variables and fertility via other explanatory variables, influencing for example fertility intentions. Others, however, play a more direct role in the explanation of fertility, facilitating, for example, certain types of behaviour such as the use of contraception.
19. Region and type of place of residence	Qs.101–103 in respondent's background	Regional and urban-rural differentials in fertility have been found almost universally. The main area for analysis is the determination of the mechanisms that bring about these differentials, through the study of the characteristics of the communities and regions and the way in which they can affect fertility.
20. Agricultural or industrial development levels	Community-level Module	A first step in the elucidation of observed regional differentials may be the consideration of development levels, a complex cluster of variables which is considered in a special module.
21. Educational levels and facilities	Community-level Module	We consider here aspects of the social structure which affect socio-cultural variables by facilitating the education of individuals and, perhaps more indirectly, by the impact of social interaction within a more educated community.
22. Health levels and facilities, including maternal and child health centres and family planning clinics	Community-level Module	Similarly, we have here aspects of the social structure which affect bio-social variables by improving general health conditions. This includes the existence of a family planning programme, which may increase the knowledge of and facilitate the use of contraception.

5. A Framework for Fertility Analysis

In our discussion of *dependent*, *intermediate* and *explanatory* variables we have provided some indication of the complex nature of the underlying system of structural relationship between the variables. This system is presented in graphical form in the figure below, which is similar to a framework proposed by Freedman (1967).
This diagram, however, is probably simpler than will be desired in particular cases and is presented for illustrative purposes only.

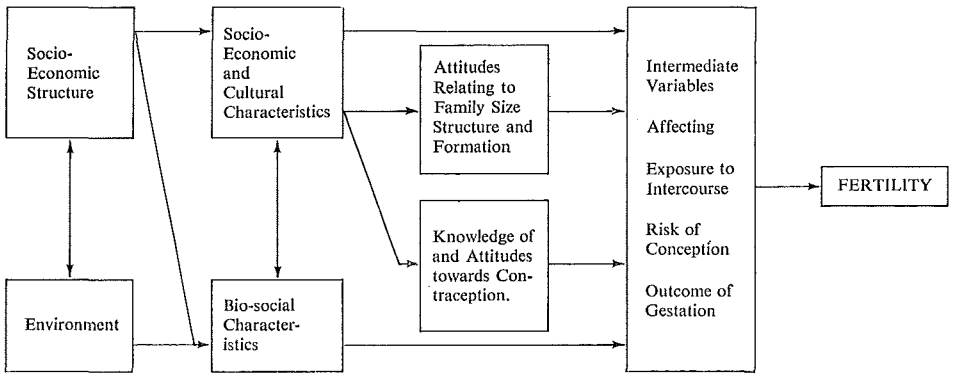
a) The relative importance of biological as against societal factors may be expected to vary very much from one country to another.

b) The framework does not necessarily show all the feedback mechanisms and linkages in the system under study. It is a model or a conspectus of hypotheses under study. Model building, especially of complicated interactive systems necessarily begins simply and works towards greater complexity until a good fit to observation is obtained.

c) Many of the causal chains connecting the explanatory to the dependent variables are obvious from general knowledge and many others have been proposed by demographers. It is, however, possible that in the light of statistical analysis some may be found to be non-contributory and new ones may be discovered. For this reason, later country recruits to the WFS will no doubt wish to study the results produced by their fore-runners.

d) As noted earlier, some of the intermediate variables and many explanatory variables are not contained in the Core Questionnaire itself but appear in the modules. The framework of analysis accordingly depends on the extent to which modular material has been collected.

A FRAMEWORK FOR FERTILITY ANALYSIS



6. Fertility Analysis and Some Lines of Inquiry

Some of the information in WFS surveys has an interest apart from questions of fertility, e.g., the educational status of different ethnic groups, or the average duration of consensual unions in different age groups. This document does not attempt to cover all the social analysis which can be carried out on the data, but concentrates on the primary interest, namely the analysis of fertility.

Even with this restriction it is clearly impossible to display all the possible lines along which the analysis of the WFS data could be conducted. The following sections are purely illustrative of the kind of approaches which are likely to be fruitful. Other suggestions on possible lines of inquiry have been given in Tables 1–3.

6.1 DESCRIPTIVE ANALYSIS AND THE MEASUREMENT OF FERTILITY

The simplest kind of analysis which can be undertaken on the WFS data is descriptive; that is to say, the information in the survey is re-arranged or re-classified so as to exhibit in a clear and summary form salient features of demographic interest. Such, for example, would be an analysis of birth intervals, of conjugal mobility, of knowledge and use of contraceptive methods, or of the distribution of duration of lactation. A great deal of the material required in this context may be explicit in the tabulations produced for Report No. 1, where the general object would be to describe (rather than account for) patterns of behaviour.

Of particular interest here is the assessment of fertility levels. The analyst has available for this purpose a large variety of measures including period and cohort age-specific and cumulative fertility rates, parity progression ratios, and indirect estimation procedures such as the "own children" technique. Where necessary, the WFS may prepare Technical Bulletins dealing with these measures.

The measures mentioned in the previous paragraph are primarily concerned with the **quantum** of fertility, i.e., the number of live births produced. Another area of interest is the **tempo** of fertility, i.e., the timing and spacing of births. Again, a variety of measures is available to the analyst, most of them based on a life-table approach, such as the mean length of birth intervals. Some of these measures might be discussed in the Technical Bulletin Series.

In addition to simple measures of fertility levels, the analyst may consider mathematical models where a probability distribution is fitted to variables such as the length of the open interval, and the parameters of the distribution are estimated from the data. There have been many models proposed in the literature; for a review of some of them see Sheps, Menken and Radick (1969).

Besides fertility itself, the analyst may be interested in the description of intermediate variables such as use of contraception, infecundity and foetal mortality. Of particular interest here is

the study of nuptiality including age at marriage, duration of marriage, marital dissolution and remarriage. The description of some of these variables may require special techniques, such as the computation of "singulate mean age at marriage", if applicable. To the extent that any such technique is found useful, it might become the subject of a Technical Bulletin.

So far, we have considered fertility and nuptiality separately; but these aspects can be integrated by recognizing that they are only stages in a complex process of family formation which involves a **sequence of events** over time. This sequence includes puberty, marriage, exposure to intercourse, conception, gestation, progression from one parity to the next, marital dissolution, remarriage and menopause. Each event is contingent upon previous ones, and at each stage reproductive choices can be made. This orientation is not only conceptually sensible, but also follows naturally from the structure of the WFS Questionnaire, where the sequence of questions follows the woman's life-cycle. The techniques of analysis are mostly of the life-table type, and produce a series of survival curves showing the passage of respondents from one stage to another.

6.2 THE ANALYSIS OF FERTILITY DIFFERENTIALS

More complicated studies would be concerned with patterns of relationship, for example between fertility and educational status, between fertility and household structure, between fertility and infant mortality, and between certain cultural factors and willingness to accept family planning. In this case also, some of the material required may exist in the form of bivariate tables in Report No. 1.

In many cases, the techniques mentioned in the previous section would be used to assess the fertility levels of different groups in the population and to show differences or relationships, usually considering only a few variables at a time. An example would be the assessment of urban-rural fertility differentials, controlling for such things as age and marital duration.

It must be recognized, however, that such simplified expositions of a complex situation must be interpreted very cautiously. The observed relation between a pair of variables may be real, permanent and significant but may not be causal. Suppose, for instance, that a relationship was found between educational status and fertility, but that educational status was highly correlated with income, and it was economic circumstances which were determining family size (through some intermediate variables). In this case, the apparent relationship between education and fertility might be due to a circuitous chain through income, and it could not be inferred without further examination that fertility could be changed by increasing the educational status of the population.

Indeed, very often the study of differentials is only the first step in a complex analytical process, where a relationship is found but the causal mechanisms involved and, in particular, the mediating variables need to be specified and studied. Thus the study of observed differentials leads logically into the building of causal models.

6.3 MODEL BUILDING IN FERTILITY ANALYSIS

Model building may start from two different points: either with a clearly formulated hypothesis which it is desired to test, or from observed relationships involving several variables, which may suggest further hypotheses for examination. On occasion, both approaches may be employed, with a transfer of emphasis from hypothesis to data and back to hypothesis, sometimes several times, in order to arrive at a satisfactory explanation of the phenomenon under study.

In all cases, it is probably best to start from a diagram of the type on page 21 setting out the variables which are to be brought into play and the measures which it is proposed to use. If possible the relationships which are to be assumed should be indicated on the diagram by lines joining the relevant variables, and if a causal connection is posited the lines should be given directional arrows. The diagram then becomes a flow chart of the dynamic movement of the system. For precise quantitative study, it will then be necessary to assume, or to estimate, the relations involved. Certain of the methodological aspects involved may be considered in the series of Technical Bulletins (see also Section 7 below).

Examples of the kind of propositions which may arise for examination are the following:

a) SOCIETIES WITH HIGH FERTILITY

i) That in the absence of socio-economic change, socio-cultural factors will be the prime determinants of high levels of fertility;

ii) That, however, such differentials as occur will result primarily from socio-cultural forces operating on factors other than contraception, in particular on abstention and duration of exposure to intercourse;

iii) That differentials in such populations may also reflect variations in environmental conditions.

b) SOCIETIES UNDERGOING RAPID FERTILITY CHANGE

i) That major differentials will be related to socio-economic factors rather than to socio-cultural factors. (These socio-economic factors will be determined primarily from the monetary rather than the subsistence sector of the economy);

ii) That differentials will be implemented primarily through fertility regulation;

iii) That, although factors other than contraception will continue to play a role in determining differentials, this role may be inversely related to contraceptive use and efficacy;

iv) That socio-economic factors will be related to differentials even when there is a control for demographic factors (particularly age), but within socio-economic groups there may be major differences between cohorts at various key periods in the life cycle.

c) SOCIETIES WITH LOW FERTILITY

i) That such things as urban residence, higher education and labour force participation are related to differentials, even if small;

ii) That cohort and other differentials result primarily from contraception and abortion, as related to changing patterns of pre-marital exposure, intercourse and conception, nuptiality and timing and spacing of pregnancies within marriage.

It will be understood that these are only a handful of propositions chosen for illustrative purposes from an enormously large number of possibilities. The types of propositions or models to be considered in any given country will depend on its specific situation and will probably be determined on the basis of the results obtained at the Report No. 1 stage or in previous fertility surveys.

7. Some Notes on Methodology

Simple patterns of relationship can often be exhibited in tabular or graphical form. But when a number of variables are to be considered simultaneously more elaborate statistical measures are required. And when, to add a further degree of complication, several different relationships are involved simultaneously, some technique has to be evolved to handle sets of questions, or contingency tables of a multivariate type.

The methods customarily used by statisticians for representing the dependence of a variate on a number of others involve regression equations. Some of the problems involved in such methods are considered in a Technical Bulletin. When the dependent variable is time to a certain event, problems of censoring must be considered; fortunately in recent years a methodology has evolved for the analysis of life-tables with concomitant information using regression-type models; this might also form the subject matter for a Technical Bulletin. Multivariate contingency or categorization, the analogue for classified data to regression or variance-analysis for continuous data, is a relatively new subject and it is possible that this subject will be raised in the Technical Bulletins.

The handling of sets of relations, which becomes necessary in more complex model building, is a subject which has engaged the attention of econometricians for some years and has only recently begun to be applied in demography. We refer here to path analysis and simultaneous equation models. Path Analysis is considered in a Technical Bulletin.[4]

Before proposing methods suitable for the analysis of fertility, the WFS intends to do some studies on a few of the early surveys to see how far existing methodology is appropriate and how far new methodology should be developed. The results of some of these studies will be published in Technical Bulletins. In the meantime the WFS staff is prepared to advise on particular problems of analysis so far as lies within their power.

REFERENCES

K. DAVIS and J. BLAKE, (1956) "Social Structure and Fertility: An Analytic Framework" *Economic Development and Cultural Change*, 4, pp. 211-235.

R. FREEDMAN, (1967) "Applications of the Behavioural Sciences to Family Planning Programs", *Studies in Family Planning*, No. 23, pp. 5-9

E. GREBENIK and A. HILL, (1974) "International Demographic Terminology: Fertility, Family Planning and Nuptiality," *IUSSP Papers*, No. 4.

M. C. SHEPS. J. A. MENKEN and A. P. RADICK (1969) Probability Models for Family Building - an Analytical Review, *Demography*, 6, pp. 161-183.

UNITED NATIONS (1958) *Multilingual Demographic Dictionary*, Population Studies No. 29.

4 M. G. Kendall and C. A. O'Muircheartaigh "Path Analysis and Model Building", Technical Bulletins Series, (The Hague. International Statistical Institute, 1977).