Scientific Reports

NUMBER 37 DECEMBER 1982

H.J. PAGE
R.J. LESTHAEGHE
I.H. SHAH

Illustrative Analysis: Breastfeeding in Pakistan

INTERNATIONAL STATISTICAL INSTITUTE
Permanent Office. Director: E. Lunenberg
428 Prinses Beatrixlaan, PO Box 950
2270 AZ Voorburg
Netherlands

WORLD FERTILITY SURVEY
Project Director:
Halvor Gille
35-37 Grosvenor Gardens
London SW1W 0BS, UK

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, Voorburg, The Hague, Netherlands.

L'Enquête Mondiale sur la Fécondité (EMF) est un programme international de recherche dont le but est d'évaluer l'état actuel de la fécondité humaine dans le monde. Afin d'atteindre cet objectif, des enquêtes par sondage sur la fécondité sont mises en oeuvre et financées dans le plus grand nombre de pays possible, Ces études, élaborées et réalisées de façon scientifique, fournissent des données représentatives au niveau national et comparables au niveau international.
L'Institut International de Statistique avec l'appui des Nations Unies, a été chargé de la réalisation de ce projet en collaboration avec l'Union Internationale pour l'Etude Scientifique de la Population. Le financement est principalement assuré par le Fonds des Nations Unies pour les Activités en matière de Population et l'Agence pour le Développement International des Etats-Unis.
Cette publication fait partie du programme de publications de l'EMF, qui comprend la Documentation de base, les Documents Non-Périodiques et des publications auxiliaires. Pour tout renseignement complémentaire, s'adresser au Bureau d'Information, Institut International de Statistique, 428 Prinses Beatrixlaan, Voorburg, La Haye, Pays-Bas.

La Encuesta Mundial de Fecundidad (EMF) es un programa internacional de investigación cuyo propósito es determinar el estado actual de la fecundidad humana en el mundo. Para lograr este objetivo, se están promoviendo y financiando encuestas de fecundidad por muestreo en el mayor número posible de países. Estas encuestas son diseñadas y realizadas científicamente, nacionalmente representativas y comparables a nivel internacional.
El proyecto está a cargo del Instituto Internacional de Estadística en cooperación con la Unión Internacional para el Estudio Científico de la Población y con la colaboración de las Naciones Unidas. Es financiado principalmente por el Fondo de las Naciones Unidas para Actividades de Población y por la Agencia para el Desarrollo Internacional de los Estados Unidos.
Esta publicación ha sido editada por el Programa de Publicaciones de la EMF, el que incluye Documentación Básica, Publicaciones Ocasionales y publicaciones auxiliares. Puede obtenerse mayor información sobre la EMF escribiendo a la Oficina de Información, Instituto Internacional de Estadística, 428 Prinses Beatrixlaan, Voorburg-La Haya, Países Bajos.

Scientific Reports

Illustrative Analysis: Breastfeeding in Pakistan

H.J. PAGE
Centrum voor Sociologie, Vrije Universiteit Brussel
Faculteit der Rechtsgeleerdheid, Rijksuniversiteit Gent
Interuniversity Programme in Demography, Brussels

R.J. LESTHAEGHE
Vrije Universiteit Brussel
Interuniversity Programme in Demography, Brussels

I.H. SHAH
At time of writing, doctoral candidate in the
Interuniversity Programme in Demography, Brussels,
on leave from Pakistan Institute of Development Economics

Typeset and printed in Great Britain by H Charlesworth & Co Ltd

Contents

TABLES

FIGURES

Preface

One of the main concerns of the World Fertility Survey has been the analysis of the data collected by the participating countries. It was decided at the outset that, in order to obtain quickly some basic results on a comparable basis, each country would produce soon after the fieldwork a First Country Report, consisting of a large number of cross-tabulations with a short accompanying text. Precise guidelines for the preparation of the tables were produced and made available to the participating countries.

It was also recognized, however, that at later stages many countries would wish to study in greater depth some of the topics covered in their first reports, or indeed new but related subjects, using more refined analytic techniques. In order to assist the countries at this stage, a general 'Strategy for the Analysis of WFS Data' was outlined, a series of Technical Bulletins was started, dealing with specific methodological issues arising in the analysis, and a list of 'Selected Topics for Further Analysis of WFS Data' was prepared, to serve as a basis for selecting research topics and assigning priorities.

It soon became evident that many of the participating countries would require assistance and more detailed guidelines for further analysis of their data. Acting upon a recommendation of its Programme Steering Committee, the WFS then launched the present series of 'Illustrative Analyses' of selected topics. The main purpose of the series is to illustrate the application of certain demographic and statistical techniques in the analysis of WFS data, thereby encouraging other researchers and other countries to undertake similar work.

In view of the potentially large number of research topics which could be undertaken, some selection was necessary. After consultation with the participating countries, 12 subjects which are believed to be of top priority and of considerable interest to the countries themselves were selected. The topics chosen for the series span the areas of fertility estimation, levels, trends and determinants, marital formation and dissolution, breastfeeding, sterilization, contraceptive use, fertility preferences, family structure, and infant and child mortality.

It was envisaged that each study would include a brief literature review summarizing important developments in the subject studied, a clear statement of the substantive and methodological approach adopted in the analysis, and a detailed illustration of the application of such an approach to the data from one of the participating countries, but with emphasis on the general applicability of the analysis. These studies have been conducted in close collaboration with the country concerned, where possible with the active participation of national staff.

It should perhaps be emphasized that the studies in the 'Illustrative Analyses' series are meant to be didactic examples rather than prescriptive models of research, and should therefore not be viewed as cookbook recipes to be followed indiscriminately. In many cases the investigators have had to choose a particular course of action from several possible, sometimes equally sound, approaches. In some instances this choice has been made more difficult by the fact that demographers or statisticians disagree among themselves as to the approach most appropriate for a particular problem. In the present series we have, quite intentionally, resisted the temptation to enter the on-going debates on all such issues. Instead, and in view of the urgency with which countries require guidelines for analysis, an attempt has been made to present what we believe to be a basically sound approach to each problem, spelling out clearly its drawbacks and limitations.

In this difficult task the WFS has been aided by an *ad hoc* advisory committee established in consultation with the International Union for the Scientific Study of Population (IUSSP) and consisting of Ansley Coale (Chairman), Mercedes Concepción, Gwendolyn Johnson-Acsádi and Henri Leridon, to whom we express our gratitude. Thanks are also due to the referees who have generously donated their time to review the manuscripts and to the consultants who have contributed to the series.

Many members of the WFS staff made valuable contributions to this project, which was co-ordinated by V.C. Chidambaram and Germán Rodríguez.

HALVOR GILLE
Project Director

Acknowledgements

The authors are very grateful to Germán Rodríguez, John Hobcraft, John McDonald, David Smith and Anrudh Jain for their helpful comments on the first draft of this report.

1 Introduction

1.1 BROAD AIMS OF THE ANALYSIS

In view of the importance of breastfeeding, both in terms of infant health and in terms of its impact on fertility, it is extraordinary that so little systematic detailed information about breastfeeding patterns throughout the world has been available until recently. As a recent review of the literature reveals (Cole 1979), much of the information that has been available to date refers to small or to peculiar samples; moreover, many analyses have used questionable or unstated methods, making interpretation of the results difficult.

Data collected in connection with the World Fertility Survey – especially in those countries that have adopted the module, Factors other than contraception affecting fertility – constitute a major step forward, by furnishing essentially comparable data for samples that are both large and representative. They thus provide the means for systematic estimation and analysis of breastfeeding distributions at the national level and also the possibility of analysing both within-country and between-country differentials.

This should not detract attention, however, from the fact that problems remain. The WFS data on breastfeeding are restricted in scope with reference to the number of births for which information was collected from each woman and also, particularly in countries that collected breastfeeding data only in the core questionnaire, in the breadth and depth of the questions asked. The development of appropriate methods for the analysis of breastfeeding data, especially data collected in single-round surveys, is relatively new. For some substantive questions, new methodological strategies are still in the process of technical development and evaluation; even for many of the more fully developed methods, practical experience of their application is still in its infancy. Thus although for certain questions it is possible to recommend a particular method of analysis as analytically sound and practical, for other questions no single 'best' strategy can yet be clearly recommended, given the present state of the art.

In this particular analysis, we therefore lay particular stress on outlining not only the potential but also the limitations of the data available, on mapping some of the major pitfalls that can trap the unwary, and on the exploratory and tentative nature of some of the analysis. Where a clear methodological recommendation can be made, we outline our reasons for choosing this method; where the situation is less clearcut we illustrate a selection of the strategies that are currently available, discussing the advantages and disadvantages of each.

1.2 ORGANIZATION AND SCOPE OF THE ANALYSIS

In order to set the framework within which evaluation, analysis and interpretation of actual data can proceed, we begin with a very brief sketch of the demographic roles of breastfeeding. We then discuss the nature of the data on breastfeeding available from WFS surveys in general, and from the Pakistan Fertility Survey in particular, in order to assess the extent to which they can respond to major substantive questions concerning breastfeeding practices and their demographic impact.

The actual analysis then opens with a discussion of the methods of using these data to estimate breastfeeding patterns; an illustration of their application is given in the estimation of national characteristics. Analysis of breastfeeding differentials follows. The analysis ends with a section addressing assessment of the impact of breastfeeding on fertility.

The mandate for this particular paper was to illustrate the analysis of breastfeeding data from the WFS core questionnaire. The special module, Factors other than contraception affecting fertility (FOTCAF), that was used in some countries, yields considerably more information about the impact of breastfeeding than the core questionnaire, since it addresses directly the principal mechanisms by which breastfeeding impinges on fertility (the post-partum anovulatory period and also, in some populations, post-partum abstinence). Although an actual illustration of the richer analyses that can be made using that module is beyond the scope of the present report, much that is said here is also applicable to that module. In particular we should note that the basic methods applied here to estimating breastfeeding patterns from the core questionnaire can also be applied to the breastfeeding data collected in the module, and to the data the latter contains on post-partum amenorrhoea and post-partum abstinence. We have attempted to draw attention wherever possible to major similarities and differences between the core questionnaire and the FOTCAF module so that the present report can serve not only as an illustration of analysis of the core questionnaire but also as a partial illustration for the module.

2 The Demographic Importance of Breastfeeding

A brief sketch of the role of breastfeeding is needed to set the context for interpretation of breastfeeding data.

2.1 BREASTFEEDING AND FERTILITY

In the context of the World Fertility Survey, primary interest in breastfeeding lies in its fertility-inhibiting effects.

Post-partum Infecundability

Following childbirth, each woman experiences a period of temporary infecundability, commonly referred to as the post-partum non-susceptible period, during which she does not ovulate. Related to this, although not necessarily lasting exactly the same number of months, is a period of amenorrhoea. Since amenorrhoea is easier to observe than anovulation, post-partum amenorrhoea is often used as a convenient operational definition of the post-partum non-susceptible period. Breastfeeding practices appear to be the principal determinant of variations in the length of this period.

In the absence of breastfeeding, post-partum amenorrhoea commonly lasts about two months on average (Potter *et al* 1965; Salber *et al* 1966; Perez *et al* 1971; Bonte and Van Balen 1969; Chen *et al* 1974). Where breastfeeding is prolonged and intensive, average post-partum amenorrhoea can last between one and two years (Chen *et al* 1974; Singarimbum and Manning 1976; Huffman *et al* 1978; Cantrelle and Ferry 1979). Breastfeeding is thus capable of increasing the average interval between successive births by up to about 18 months. Expressed another way, if the average interbirth interval in the absence of breastfeeding were to be, say, 20 months for women in the central reproductive period (two months 'minimum' mean post-partum amenorrhoea, nine months for waiting time to conception and any time lost through pregnancy wastage, and nine months gestation), then universal, prolonged and intensive breastfeeding could almost double the interbirth interval, nearly halving the fertility rate for these women. Conversely, complete disappearance of an existing pattern of prolonged and intensive breastfeeding could lead to an almost doubling of their fertility (unless compensated by adoption of other practices that inhibit fertility). Clearly, through their suppression of ovulation and menstruation breastfeeding patterns and trends can play a major role in determining fertility levels and trends.

Although the relationship between duration of breastfeeding and duration of post-partum amenorrhoea is very striking, it is not perfect. Some women who breastfeed for a long time experience relatively short post-partum amenorrhoea, and vice versa. Furthermore, two populations with the same average reported duration of breastfeeding may have rather different average durations of amenorrhoea. Thus while a one-month difference between two populations in their mean duration of breastfeeding corresponds, on average, with about a one-half month difference in their median duration of amenorrhoea (Leridon 1977; Santow 1978; Corsini 1979) – perhaps with rather more in the central range of values and rather less at very low or very high durations (Lesthaeghe and Page 1980; Bongaarts 1981) – populations exhibit a fair amount of scatter about the general curve. A number of factors may contribute to this scatter, but a major probable contributor is the fact that two groups of women with the same overall duration of breastfeeding may have quite different patterns with respect to their frequency and intensity of breastfeeding. Women who are giving full (unsupplemented) breastfeeding, for example, have lower chances of having resumed menstruation than women who are giving their children supplementary foods (Perez *et al* 1971; Huffman *et al* 1978).

The principal explanation lies in the endocrine factors that are associated with lactation.[1] Lactation itself depends on the secretion of the hormone prolactin by the anterior pituitary. At delivery, a woman's prolactin levels are high. In the absence of breastfeeding, serum prolactin concentrations tend to decline to pre-pregnancy levels within about a week; high levels are usually maintained, however, if the child is breastfed. Each time the infant suckles, the stimulus of the breast-nipple triggers a neurally mediated hormonal reflex by which more prolactin is promptly secreted, leading to a temporary elevation of serum prolactin levels. Although the exact mechanisms operating are not entirely clear, it is apparent that the high prolactin levels not only maintain milk production but also are associated with an inhibiting effect on hormones that regulate ovulation and the menstrual cycle.[2] Frequent and relatively intense suckling may be needed, however, to maintain this inhibition. It appears that prolactin concentrations tend to vary with the intensity of breastfeeding, being higher among women who are fully breastfeeding than among those who are only partially breastfeeding (Delvoye *et al* 1976 and 1977; Tyson *et al* 1976). Less frequent suckling results in less frequent prolactin elevations and lower serum levels of ovarian hormones (Delvoye *et al* 1976 and 1977; Tyson and Perez 1978; Konner and Worthman 1980; Howie 1981). Thus the effect of lactation on post-partum anovulation and amenorrhoea may well depend on the frequency and intensity of breastfeeding. An overall breastfeeding duration of 12 months where the transition from full and frequent

[1] See, for example, Tyson and Perez (1978), McNeilly (1979) and Howie and McNeilly (1982b) for a recent summary.

[2] High prolactin levels are known to be associated with inhibition of gonadotrophin release from the pituitary (Tyson *et al* 1976). They also appear to inhibit ovarian steroid synthesis directly or reduce the sensitivity of the ovary to pituitary gonadotrophin stimulation (Bonnar *et al* 1975; Rolland *et al* 1975 for example).

breastfeeding to none at all occurs swiftly at the end of the period may have very different implications for anovulation and amenorrhoea than the same overall duration where the infant generally suckles less frequently and where supplements are introduced very early.

It has also been suggested that even among women who have resumed menstruation, those who are breastfeeding may have slightly lower probabilities of conception than non-nursing women, and hence longer average waiting times to conception. This could occur if, for example, breastfeeding women have a higher chance of anovulatory menstrual cycles, of less favourable luteal phases, or of interference with implantation. However, evidence of a significant delay in conception after resumption of menstruation (Jain *et al* 1979; Howie and McNeilly 1982a) is much more meagre than the evidence that breastfeeding has a major impact on amenorrhoea.

We can summarize this discussion in three short points:

1 the principal effect of breastfeeding on fertility among most populations operates through its physiological impact on post-partum anovulation and amenorrhoea.
2 breastfeeding appears to be the main determinant of variations between populations in their mean duration of post-partum amenorrhoea.
3 the exact relationship between the duration of breastfeeding and that of post-partum amenorrhoea is not everywhere the same; the relationship probably depends quite heavily on the frequency and intensity of breastfeeding, although other factors (such as nutrition) may also play a role.

The implications for estimating the physiological impact of breastfeeding on fertility are obvious. We really need direct information on the length of the post-partum non-susceptible period as well as on breastfeeding. Although data on anovulation would be best, we can probably work fairly adequately with data on amenorrhoea, which are easier to collect. Ideally, we would also want data on the frequency and intensity of breastfeeding, not just its overall duration, in order to evaluate more fully the relationship between breastfeeding practices and amenorrhoea.

If no data on amenorrhoea are available we have only two possibilities: either we must assume that the population in question lies on a curve expressing the 'average' relationship observed between breastfeeding and amenorrhoea (and run the risk that the average relationship is not a good predictor for this population) or try to estimate the impact indirectly by relating the duration of breastfeeding to the length of the interbirth interval or to some other measure of fertility (in which case we may not easily be able to separate out the effects of breastfeeding from those of other factors that can affect fertility). Clearly actual data on amenorrhoea are desirable, and it is unfortunate that the WFS core questionnaire did not include any questions on this.

Post-partum Abstinence

In addition to its physiological impact, breastfeeding can also have a socially mediated impact on fertility. In a number of populations, most notably in tropical Africa, sexual intercourse is traditionally either totally proscribed or at least restricted for the new mother for a period that may vary from several months to several years after each delivery (Schoenmaeckers *et al* 1981). The taboo is frequently related to breastfeeding, sexual activity being proscribed all the time the child is heavily dependent on breast-milk, or throughout the entire period of breastfeeding, or even – as among the Yoruba, for example (Caldwell and Caldwell 1977) – for a period that extends for several months after the child is fully weaned. It would be an oversimplification to think of this practice simply as a 'lactation taboo' for the association is not perfect and there are a number of other cultural and social factors that underlie post-partum abstinence traditions (Caldwell and Caldwell 1981; Lesthaeghe *et al* 1981); but nevertheless post-partum abstinence is frequently quite clearly related to breastfeeding. Of the functions that abstinence fulfils, those that are most readily perceived by the individuals concerned and most frequently cited as reasons for observing abstinence centre on the maintence of child health and particularly on the relationship between breastfeeding and child health: indeed, it is true that a new pregnancy will jeopardize the milk supply for the existing child and hence perhaps the child's health. In addition, the beliefs that ensure observance of the taboo frequently centre on breastfeeding and child health (for example, the belief that semen will actually enter the breast-milk or otherwise impair it, thus poisoning the child). Thus breastfeeding constitutes both one of the most manifest functions of post-partum abstinence and also one of the mechanisms by which observance of abstinence is ensured.

Where strong traditions of post-partum abstinence exist, the period of abstinence may well exceed the post-partum anovulatory period. Thus in some societies, post-partum abstinence is an even stronger inhibitor of fertility than post-partum amenorrhoea. Where abstinence traditions are combined with very long breastfeeding, abstinence may last two or even three years on average after each birth, leading to average birth intervals of three or four years or more. Again the WFS core questionnaire contains no specific questions on post-partum abstinence durations. And since the relationship between the duration of abstinence and breastfeeding is much more variable than that between breastfeeding and amenorrhoea, there is no way of establishing an 'average' relationship across several populations that could be used to estimate the duration of abstinence from the data on the duration of breastfeeding. Fortunately, many of the countries in which abstinence is known to be widespread adopted the module on factors other than contraception affecting fertility, which did include some questions on abstinence.

2.2 BREASTFEEDING AND INFANT HEALTH AND SURVIVAL

We have already mentioned in passing the relationship between breastfeeding and child health, and there is no doubt that breastfeeding can play a very important role in child health. Breast-milk usually meets all the child's nutritional requirements, both in quantity and composition, for the first few months of life; even later, after 4–6 months, as supplementary foodstuffs become increasingly needed, it can still meet a substantial part of the child's requirements. Few other diets are so well suited both to an infant's nutritional needs and to its digestive system. The child's health may also benefit from additional advantages conferred by breast-milk over and above its nutritional qualities. Firstly, unlike other foods that are often inadequately prepared and sterilized, breast-milk is uncontaminated and breastfeeding therefore reduces the child's risk of ingesting harmful agents such as those responsible for many potentially serious gastro-intestinal crises among infants. Secondly, breast-milk – unlike most artificial infant foods – contains cellular components capable of ingesting potentially harmful bacteria and also some useful bacteriostatic compounds, thereby actively combating potential bacterial infections. Moreover, the milk may perhaps transfer to the child, via the antibodies it contains, some of the immunity to infection already acquired by the mother, thereby providing it with defensive reinforcements against infection at a stage when the immunities it acquired *in utero* are declining and it has still not had much time to build up its own resistance.

The World Fertility Survey's interest did not lie in child health as such – indeed a very different type of survey would have to be mounted to examine the relationships between breastfeeding and child health. These relationships are of interest within the context of WFS only inasmuch as they affect child mortality which may then, in turn, impinge on fertility. However, the possibilities for analysing the link between breastfeeding and infant mortality with the data available are severely restricted.

3 The Data on Breastfeeding in WFS Data Sets

3.1 THE DEFINITION OF BREASTFEEDING

As in many surveys, the main questionnaire used in WFS did not make very explicit exactly what was intended by the word 'breastfeeding'. The basic questions used in the English-language version of the core questionnaire were simply:

(i) 'Did you feed at the breast?'
If 'yes',
(ii) 'For how many months did you breastfeed?'

No indication is given in the questions themselves of what constitutes breastfeeding, in particular whether partial or infrequent breastfeeding is to be included as well as full breastfeeding. The key issue these questions were intended to elucidate is the duration of breastfeeding. Measurement of all duration variables presupposes that both beginning and end-point can be clearly defined. We can for practical purposes simply assume that the period of breastfeeding starts at birth, but its end-point is often less clear cut. Breastfeeding often tapers off, with its frequency and intensity declining gradually over time.[3] If no particular frequency and intensity is specified as defining the end-point of breastfeeding, then individuals are left to make their own interpretation and definition: where breastfeeding tapers off gradually, variations in interpretation might lead to variations of several months in the reported duration of breastfeeding. It seems probable, however, that most respondents interpreted the question to mean any form of breastfeeding including partial breastfeeding, as was apparently intended, although it is not clear how women reported situations where a child received just occasional suckles over a long period of time.[4]

Countries adopting the module on factors other than contraception affecting fertility had the opportunity of being more specific. The central question was rephrased slightly ('For how many months altogether did you breastfeed him/her?') and the guidelines for Interviewer's Instructions stated explicitly that the question referred to partial as well as to full breastfeeding ('What is wanted is the total time a woman breastfed her child. If a woman breastfed the child, say, only at night for the last six months of breastfeeding, then these six months should be included in the total'). Moreover, an additional question was included to check that the respondent had indeed interpreted the question this way ('After _ months you had completely stopped breastfeeding your child even once a day?')[5] Furthermore, the module provided the option of a question to assess the duration of full (unsupplemented) breastfeeding: 'How many months old was the child when you began giving him additional food along with breastfeeding?'[6] The answers to this question might be extremely useful, but again they need to be interpreted in the light of local weaning practices and the interpretation given to the expression 'additional food'.[7] Clearly the local significance of the exact terms used in any translation may have played a role here.[8]

Overall it seems reasonable to assume that most women interpreted the question as referring to the total period of breastfeeding, including partial breastfeeding. But we need to bear in mind that there may have been some slippage or some difficulty in defining the point at which breastfeeding ends. Small differences in reported durations should be interpreted cautiously; they could well be the result of different interpretations rather than the result of differences in breastfeeding behaviour.

---

[3] The end points of post-partum amenorrhoea and post-partum abstinence may also be indistinct although for slightly different reasons (Page and Lesthaeghe, eds, 1981). Unlike breastfeeding, amenorrhoea usually ends fairly abruptly, although some women may experience irregular cycles (or be unable to distinguish between episodes of post-partum bleeding and relatively early resumption of menstruation). Similarly, in populations with very strong traditions of post-partum abstinence, resumption of regular sexual activity is usually not spread out over several months. In this respect, the end-point for both amenorrhoea and abstinence is often easier to define (though not necessarily easier to recall) than that of breastfeeding. In another respect, however, amenorrhoea and abstinence are more difficult. Breastfeeding must ultimately come to an end, for the mother must at some point wean her child. Amenorrhoea and abstinence, however, can continue indefinitely if post-partum amenorrhoea shades into the amenorrhoea of menopause and post-partum abstinence into 'terminal abstinence'. Whether *post-partum* amenorrhoea and abstinence have ended may thus not always be clear, especially for older women.

[4] We should note here the quite common practice of letting the child suckle occasionally more to pacify it than to feed it, which can occur even with a child several years of age. Some societies make a clear distinction: the Havu of Kivu (Zaïre), for example, distinguish clearly between 'little suckles' (given on the spot and serving primarily to calm the child) and 'large suckles' (the principal feeds, for which the woman seeks a quiet shady spot) (Caraël 1981). Both English and French-language versions of the questionnaire emphasize the feeding aspect of suckling (using the terms 'breastfeeding' and 'allaiter au sein'), though there is no guarantee that individual translations and interpretations did so.

[5] A probe that was given a slightly different, perhaps more neutral, turn in the French-language version ('Après _ mois aviez-vous définitivement arrêté d'allaiter votre enfant, pas même une seule fois par jour?').

[6] The question was perhaps more specific (though also perhaps more westernized and oriented towards relatively rigid feeding schedules) in the French-language questionnaire: 'Combien de mois avait l'enfant lorsque vous avez commencé à lui donner un repas par jour tout en continuant d'allaiter?'.

[7] Probes following rather similar questions in western Nigerian survey work, for example, have shown that not all women necessarily interpret 'other foods' in the same way: most women there were found to have referred to the point at which they introduced traditional weaning paps (or patent baby foods), but a few referred to the giving of anything at all other than the mother's milk (sugared water, or infusions of leaves or bark).

[8] Ware (1977, p39), for example, has already drawn attention to the fact that in some languages the expression for breastfeeding may simply be 'to give milk'.

3.2 BIRTHS FOR WHICH BREASTFEEDING DATA ARE AVAILABLE

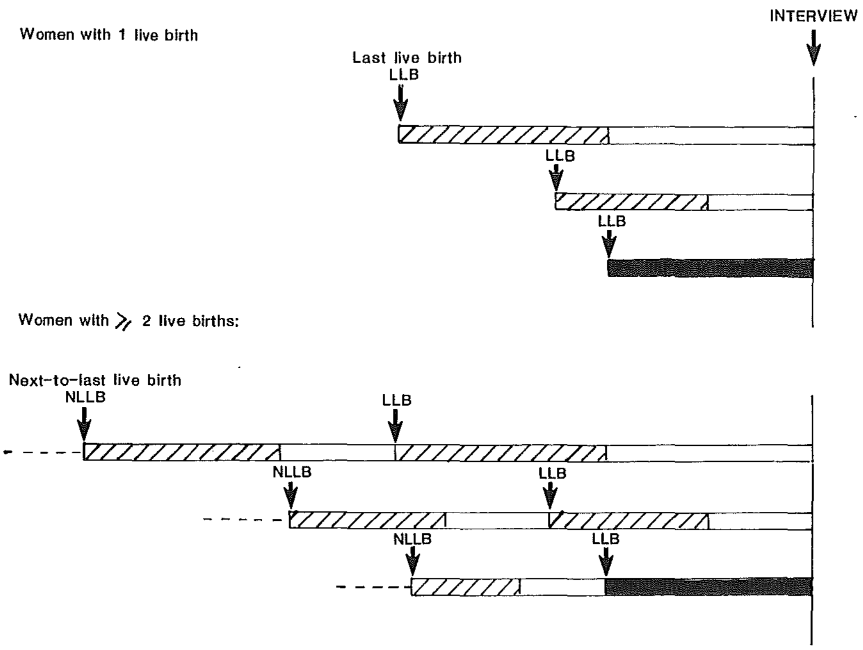
Direct Questions about Breastfeeding

The general strategy adopted in the WFS for all three post-partum variables was to restrict direct questions to each woman's two most recent live births.[9] Thus in the core questionnaire, for example, the breastfeeding questions were asked as follows (figure 1, upper section):

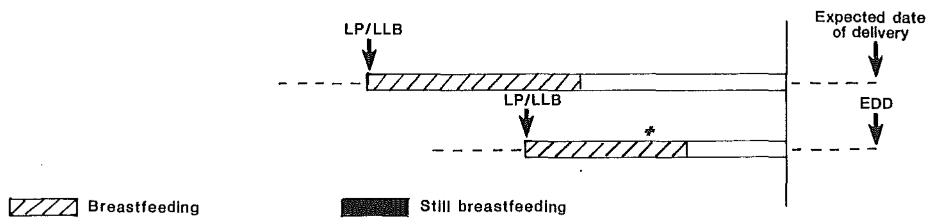
1 For women with one live birth:
- breastfeeding following the woman's most recent live birth (often called, for brevity, her 'last' live birth (LLB) or 'latest' live birth). This can also be referred to as breastfeeding in her current open birth interval (OBI).

Figure 1 Births for which direct questions on breastfeeding were included in most WFS surveys

A Countries using core questionnaire
(i) All women, regardless of reported pregnancy status at time of interview



B Countries using module, factors other than contraception affecting fertility*
(i) Women not reported as currently pregnant:
as in core questionnaire*
(ii) Women reported as currently pregnant:
All women, regardless of number of live births*



*In most countries using the module, the questions related to the two most recent pregnancies, not the two most recent live births.

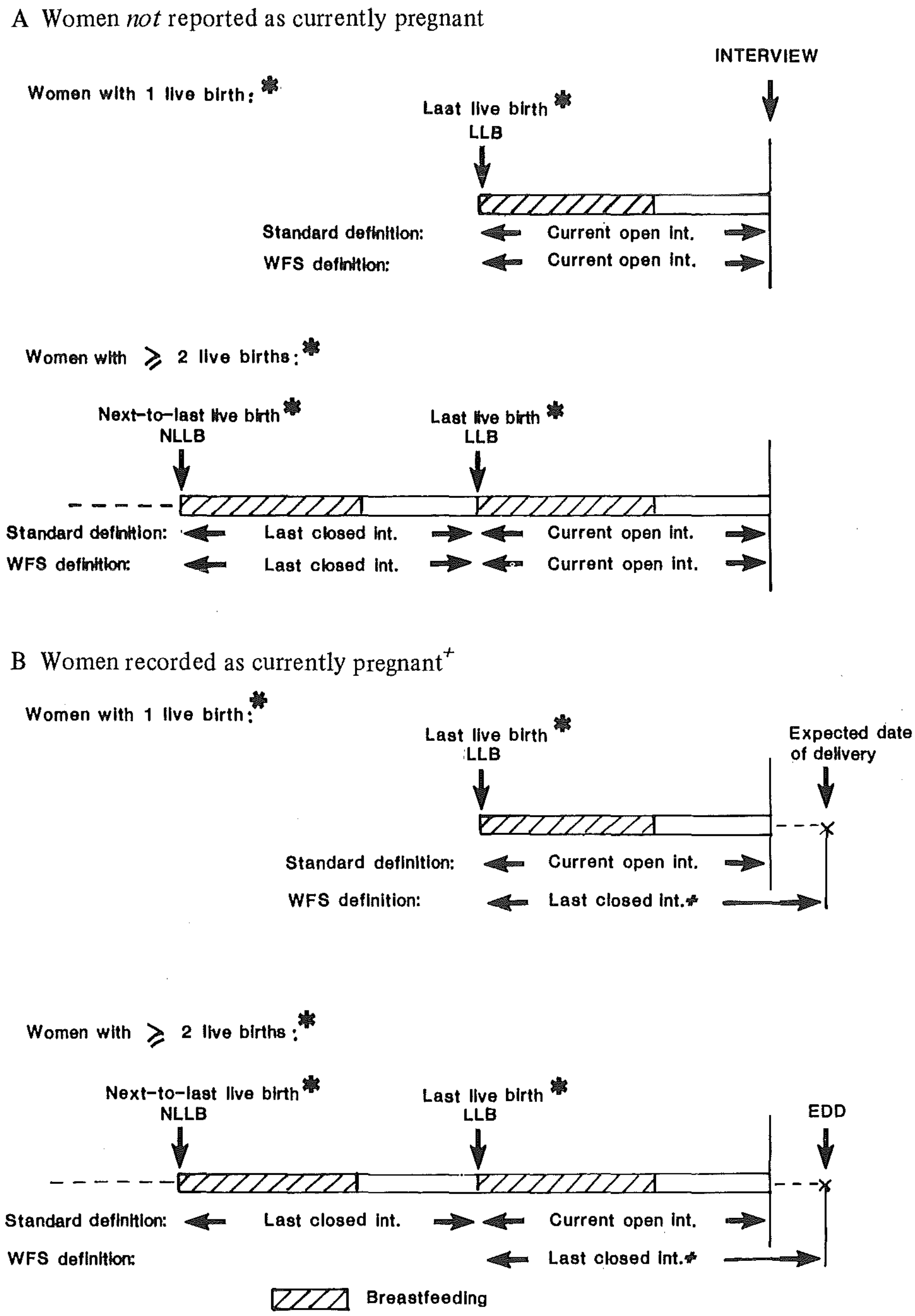
+If the woman was pregnant, no special provision was made for a response 'still breastfeeding' (it was effectively assumed that women reported as pregnant would no longer be breastfeeding): the duration of breastfeeding had to be given in months.

[9] More strictly, although the original version of the core questionnaire referred to the two most recent live *births*, the modified version used in practically all countries referred to the two most recent *confinements* that resulted in at least one live birth each. In the case of multiple births, the breastfeeding questions were to refer to the first-born of the children (or to whichever survived the longer/longest if the first-born had died). For brevity, we shall simply follow the established, albeit inexact, practice of referring throughout to 'live births' although strictly speaking we should say 'confinements resulting in one or more live births'.

2 For women with two or more live births:
- the questions were posed as above for the most recent birth;
- the same questions were also posed about breastfeeding following the preceding live birth (her 'next to last' live birth (NLLB) or 'next to latest' live birth). This can also be referred to as breastfeeding in the most recent or 'last'/'latest' closed interbirth interval (LBI).

In most countries using the module on factors other than contraception, direct questions on the post-partum variables were asked for the two most recent pregnancies, not the two most recent births. Moreover, for women who were reported as being currently pregnant, the questionnaire provided the possibility of asking these questions *only* for the most recent pregnancy and not for any preceding pregnancy (figure 1 on p12, lower section).

Figure 2 Definitions of current 'open' birth interval and of 'last closed' birth interval: comparison of standard definitions and of WFS coding conventions



*In most countries using the module, factors other than contraception affecting fertility, the data refer to pregnancies rather than to live births.

+In WFS coding conventions, women reported as pregnant at interview were defined as being currently in a closed interval, rather than an open one.

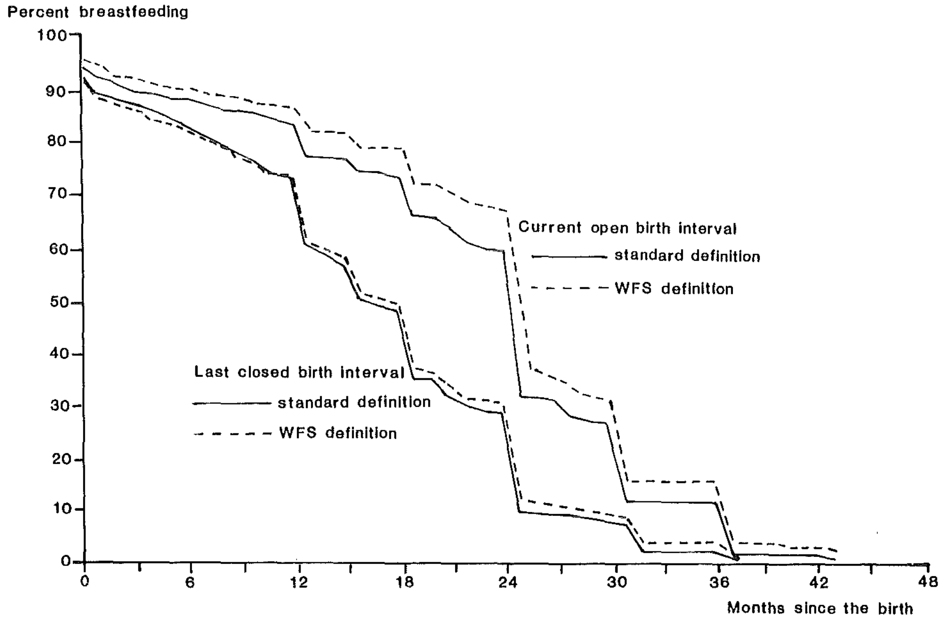


Figure 3 Comparison of estimated proportions breastfeeding derived from the two most recent births per woman using standard definitions and WFS definitions (PFS)

This restriction in the module was related to a decision taken also for the early data processing and analyses of the core questionnaire (see, for example, the First Country Reports and the Cross National Summaries volume on breastfeeding (Ferry 1981)). The decision in question concerned a rather special definition of what constituted a current open birth interval and what constituted the last closed birth interval, and the creation of a separate set of variables for each. For non-pregnant women, the standard definitions were used. For a pregnant women, however, her current open birth interval was treated as if it were her most recent closed birth interval (the interval being considered closed at the expected date of delivery). With this definition, information collected in the core questionnaire about breastfeeding following her preceding live birth (if any) was thus ignored in tabulations and analyses of the last closed birth interval; information for the most recent birth (LLB) was coded under the variables for the last closed birth interval (LCBI) rather than under the open birth interval (OBI); and no information at all about breastfeeding was coded under the variables for the open birth interval (OBI) (figure 2). In other words, pregnant women were excluded completely from the coded open birth interval data set; they were included in the closed interval data set on the strength of their most recent birth rather than the preceding one (ie they were included even if they had had only one live birth to date).

There are several analytical implications of this coding convention. The most serious concerns any attempt to analyse the length of the open birth interval as such, or of any of its components or determinants such as the post-partum variables and contraception. Since the probability of being pregnant at the time of the survey, and hence of being excluded entirely from the open interval data set, is higher for women with short interbirth intervals, these women are under-represented. To the extent that breastfeeding (or any other variable under consideration) is positively correlated with the length of a woman's birth interval, short durations of breastfeeding (or related variables) will also be under-represented.[10] Since about 10 per cent of ever-married women were typically recorded as currently pregnant in the WFS studies (the figure reaching over 20 per cent in Jordan), the effect on the data sets may be considerable. Satisfactory analysis of the length of the open interval or of its components or determinants can only be carried out if these women are included in the analysis.[11] Figure 3 illustrates how the results would be biased in Pakistan if one used the WFS definition of the open birth interval rather than the standard definition. With some data sets, the average duration of breastfeeding would be overestimated by about six months (Page *et al* 1980)! For other reasons, discussed below in chapter 4, we do not recommend use of the data from the current open birth interval for most purposes. In the few instances where we do use them in this paper, we shall use the standard definitions.

The consequences of the definitions for analysis of the last closed birth interval are more complex, but usually of less practical significance. The main reason is that currently pregnant women are not systematically excluded either by

[10] A negative correlation between breastfeeding and birth interval length is less common but can occur (for example, where those who breastfeed the least tend to have longer than average birth intervals because of greater use of contraception). In such cases, it would be the women with longer durations of breastfeeding rather than those with short intervals who would be under-represented.

[11] Fortunately, their inclusion can easily be accomplished: one simply has to assign to the open birth interval variables for pregnant women the corresponding values for their most recent birth – values that are already coded as the variables for the last closed birth interval.

the special definition or by the standard definition.[12] The two definitions are conceptually distinct, however, and will not necessarily yield exactly the same results. For example, with the WFS definition of the last closed birth interval, pregnant women are included even if they had had only one live birth, whereas in the standard definition such women are excluded. One of the potential advantages of the WFS definition is clearly that fewer women are excluded from the data set, a factor that both increases sample size and is likely to reduce selection biases (discussed in more detail in the following section). A disadvantage, however, is that the definition is not entirely 'clean'. Some of the women who are pregnant will not proceed to a live birth from their current pregnancy: they will either miscarry or experience a stillbirth. In other words, their actual interbirth interval will be longer than the interval estimated from the expected date of delivery. While this may not directly affect estimates of breastfeeding in the closed interval, it could affect not only estimated birth intervals as such but also analysis of the relationship between breastfeeding and birth interval length.

Only the special definition of the last closed interval, embodying 'artificially' closed intervals for pregnant women, can be used in countries that adopted the module on factors other than contraception, since that questionnaire never included questions on post-partum variables in the genuine closed birth interval for these women. In countries where the breastfeeding data come from the core questionnaire, however, one often has a choice as to which definition to use: if the 'genuine' last closed birth intervals are desired, the original data pertaining to the next to last live birth for each pregnant woman can usually be located and used instead of the value recorded. Again, at the points where we use the last closed interval in this paper, we shall use the standard definition.

Finally, we should point out that a few countries (of which Pakistan is one) collected breastfeeding data for *all* live births each woman reported, not just for the most recent ones. While it is unlikely that the reporting of breastfeeding durations for children born in the more distant past is accurate enough for such data to reveal any but major trends over time, the decision not to restrict data to the two most recent births does carry with it some analytical advantages. In particular it allows us to escape from some of the analytical problems (discussed in chapters 4–6) that are inherent with data sets that are restricted not just to recent births but to the two most recent births for each woman.

Some Implications of Restricting Data Sets to the Two Most Recent Births

A number of important limitations arise when data are restricted to the two most recent births. They derive from the fact that the data refer not only to a relatively short segment of each woman's experience but also to a segment that is not the same for all women.

Some of the difficulties reside in the problems common to all retrospective fertility data obtained from single-round surveys of women who are encountered at different points in their potential reproductive age-span: only partial histories can be collected for most women, the date of survey acting as a cut-off point. Older women, who are nearly at the end of their reproductive period, can give fairly complete histories; younger, and recently married, women may have barely begun and no information is available on what they will do subsequently. Thus if we want to analyse birth intervals and their components we may face large selection effects. For example, if we look at the interval between second and third births, younger or more recently married women have had less chance to reach the starting point of this interval; moreover, even within a given birth and marriage cohort, women with long interbirth intervals are less likely to have reached this interval than are women with short birth intervals. These selection effects can be handled partially through introduction of appropriate controls (such as age, marriage duration, parity), as is discussed more fully in the illustrative analysis of birth intervals (Rodríguez and Hobcraft 1980). The selection biases can rarely be entirely eliminated, however, so care must always be exercised in interpreting results by consideration of exactly which births have been included.

Restriction of data to the two most recent births introduces an additional set of restraints, however, over and above this general one. First, any analysis then excludes not only the later part of most cohorts' experience (a particularly large part for younger women); it also excludes the earlier part of most cohorts' experience (a particularly large part for older women). If one has a full maternity history up to the date of interview, one at least has information on the earliest stages of family formation for nearly all cohorts, even if the later stages are missing for younger cohorts: when data are restricted to the two most recent births, one does not even have this. Some women will be reporting only their first two births (experienced perhaps in their teens) while others will be reporting on only, say, their sixth and seventh births (experienced perhaps in their late 30s or early 40s). Secondly, using the two most recent births does not even define the same point in historical time for all women (as is illustrated in figure 1). Younger women who are still in the process of family formation may have had their most recent two births in the 3–4 years preceding the survey; older women who ceased childbearing several years before the survey may be reporting on births that occurred 10–15 years before the survey. Even more insidious is the fact that even among women who are still actively bearing children, women with short interbirth intervals will be reporting on births that occurred more recently than women with long interbirth intervals. In short, the births do not refer to experience within the same period of time. Only if we restrict ourselves to the subset of births that occurred in a very short time period immediately preceding the survey – short enough so that no woman could have had more than two births (say two years) – is the data set going to be perfectly representative of all births that occurred in the given period to the women interviewed. For any other periods, short interbirth intervals are underrepresented because some women will have had another birth during the period as well as the one(s) they reported on.

In order to obtain a representative sample of births in a

[12] Although analysts sometimes choose to exclude them (Jain and Bongaarts 1981; and Smith 1980, for example) because of incomplete data or uncertainty concerning the way in which they were interviewed or their information coded.

given time period using data restricted to the two most recent births, the length of this period must thus be rather severely restricted. This has the effect of considerably reducing the number of observations (and also of totally excluding women who ceased childbearing several years ago). Moreover, a period lasting only 1–2 years is insufficient to permit good estimates of breastfeeding durations in populations where a sizeable proportion of women breastfeed for longer than this.

Any attempt to go much further back in time using these data, or to measure trends over time, must be ruled out. The further back one goes in time, the more short birth intervals (usually those with short breastfeeding or no contraceptive practice) are under-represented. To sum up, it is impossible with these data to perform analyses for any given *time period* except for a very short period just before the survey, and it is quite impossible to measure trends over time.

Given the difficulty both of recalling events in the distant past and also of expanding a questionnaire to include questions about every birth, the decision to focus attention on recent events is entirely understandable. Good estimates of recent breastfeeding patterns will in themselves be of great value. As the following chapters show, however, restriction of the data to the last closed birth interval and the current open birth interval for each woman means that neither the estimation of recent breastfeeding patterns nor evaluation of the relationship between breastfeeding and fertility are entirely straightforward because of the selection effects. The problems are discussed in chapter 4–6. One possible way out of the apparent impasse is the use of indirect data.

Use of Indirect Data

Although no direct questions about the post-partum variables were asked for earlier births in most WFS surveys, we can make plausible assumptions. In most populations it is not unreasonable to assume that any child who was born before the mother's most recent live birth has already been weaned, since women rarely continue to breastfeed an existing child right through a full pregnancy.[13,14] We do not know how long breastfeeding actually lasted for these children but we can infer that it had ceased before the survey. In other words, we either know (from the direct questions) or can infer (in other cases) *the current breastfeeding status* of all children. Since the dates of all births were collected in the maternity history, we can use the current status data to estimate durations of breastfeeding. For example, we can easily calculate for *all* children born in a given month the proportion still being breastfed at the time of the survey: we simply combine the direct information on current breastfeeding status that was obtained for children who figured among the two most recent births with an assumption that any children who preceded them have already been weaned. The series of proportions still being breastfed thus derived represents *all* births in a given period.

We cannot use this strategy for the distant past, because it is meaningful only for a period immediately preceding the survey that does not itself exceed the longest duration of breastfeeding. For example, if we know that all children are weaned by four years of age, the proportions still being breastfed among children born more than 48 months before the survey are non-informative for they are automatically zero. However, the period that can be covered by the combination of direct questions and inferred information on current breastfeeding status may well be longer than the one or two years for which the direct data on the two most recent births are representative. This data type may therefore be particularly useful where breastfeeding is prolonged and the period that can be represented by the direct data alone is too short.

3.3 THE QUALITY OF REPORTING

The two main types of data available for use either separately or in combination – retrospectively reported durations of breastfeeding on the one hand and current breastfeeding status combined with the dates of the births concerned, on the other – are subject to rather different types of reporting errors.

Reporting of Current Breastfeeding Status

Both types of data could be affected – although not in exactly the same way – if women who were still breastfeeding their most recent child were to misunderstand the intention of the question, 'For how many months did you breastfeed?' and tried to respond in terms of an actual number of months rather than saying that they were still breastfeeding.[15] Such a misunderstanding, if it occurred, could lead to a downwards bias in the apparent proportions still breastfeeding, while its impact on retrospectively reported durations would be either downwards (if the women tended to state the number of months of breastfeeding to date) or negligible (if they tended to state the number of months they *intended* to breastfeed that child). It is hard to assess the extent to which this misunderstanding may have occurred, but at first sight at least, it appears to have been relatively rare. Usually only a few women (well under 5 per cent) report breastfeeding durations equal to or in excess of the recorded age of the child in question, and even for the latter group it may be the age of the child rather than the breastfeeding duration that is the source of the inconsistency, especially if the date of birth had to be imputed.[16]

---

[13] We can make similar assumptions about post-partum amenorrhoea and abstinence. The woman must, by definition, have resumed ovulation (even if she did not actually resume menstruation) between the two births, and she must also have resumed sexual activity.

[14] It was in fact assumed in the questionnaire that women who had closed the birth interval in question (or even, in the module, women reported as currently pregnant) had already weaned the child; and 'still breastfeeding' the child was simply not included as one of the possible responses.

[15] As Knodel and Debavalya (1980) point out, 'still breastfeeding' was not explicitly mentioned or suggested in the *question* itself as a possible response, although it was included in the printed questionnaire in the space where the interviewer noted the response.

[16] In Thailand, Knodel and Debavalya found 1–4 per cent reporting breastfeeding durations in excess of the recorded age of the child. Examination of the individual questionnaires concerned showed that none of the inconsistencies involved imputed birth dates, and their source may thus have been statement of the intended duration of breastfeeding. In Pakistan, the corresponding figure is <1 per cent. Without access to the original questionnaires, we cannot say whether or not imputation of birth dates may be the cause.

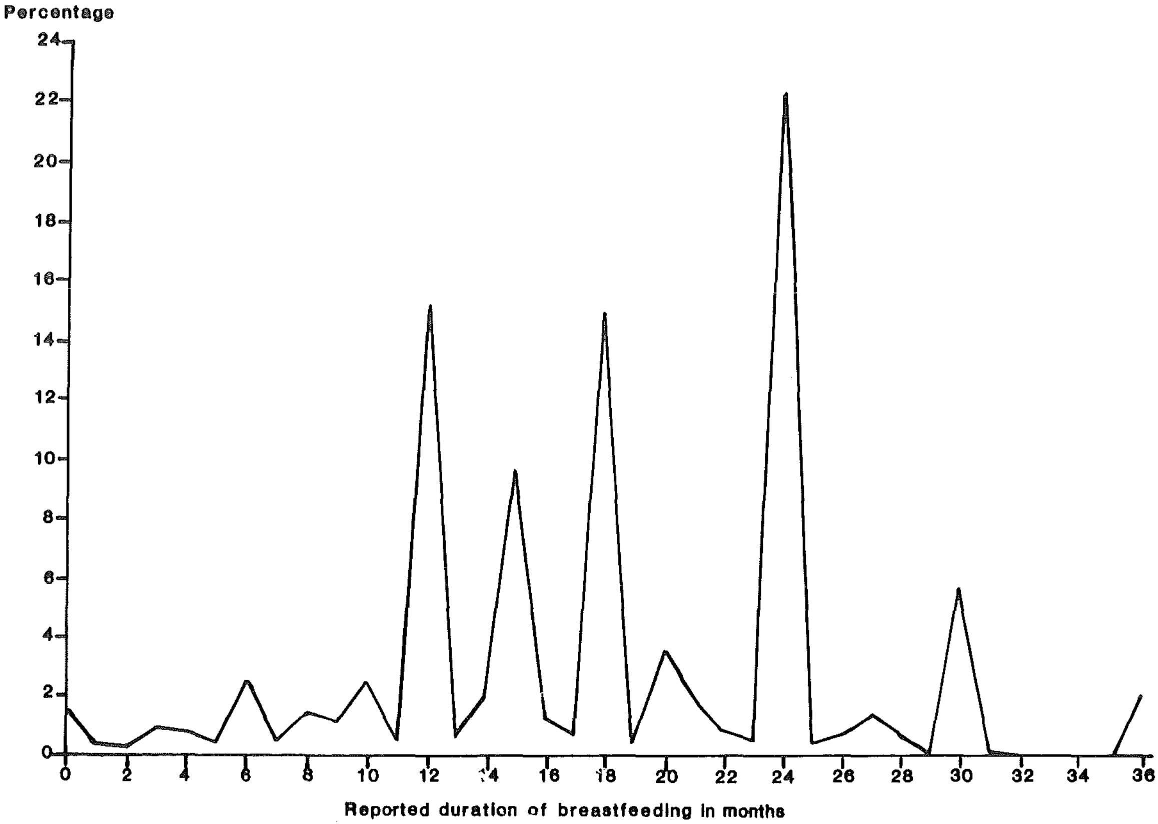


Figure 4 Retrospectively reported breastfeeding durations: percentage distribution of women by reported duration of breastfeeding following their 'next to last' live birth (PFS)

More attention has, therefore, been focussed on the effect of errors in the reported breastfeeding durations for those children who *have* been weaned or in the recorded dates of the births, neither of which are necessarily recalled with great accuracy.

Retrospectively Reported Breastfeeding Durations for Children who Have Been Weaned

With retrospective reports it is very hard to tell whether there is any systematic tendency to overstate or to understate the durations unless the errors are gross.[17]

Like many retrospectively reported duration variables, however, the frequency distributions for breastfeeding typically exhibit dramatic peaks at multiples of 6 or 12 months (Ferry 1981, figures 1–3), which arouse doubts in the minds of many sceptical analysts. In Pakistan for example (figure 4), over 50 per cent of all women reporting on their next to last live birth reported breastfeeding durations of exactly 12, 18 or 24 months. Similar patterns are found in retrospective reports of the other post-partum variables, amenorrhoea and abstinence (Lesthaeghe and Page 1980). In the case of amenorrhoea, such peaking must be almost entirely the result of rounding in reports. For breastfeeding, however, and also for post-partum abstinence, genuine peaking could occur with real normative concentrations at, or at least around, points such as 12 or 24 months. Although some researchers argue that for certain societies much of the heaping is genuine (for example, see the discussion by the Caldwells, Santow and Bracher in Page and Lesthaeghe, eds, 1981, on post-partum abstinence among the Yoruba), most demographers feel that much of the heaping is an artefact of rounding; how much is still a subject of discussion. In other words, the extent to which the shape of the distribution (also perhaps, the reported median and quartiles) may be distorted is not known. However, if there is no greater tendency to round up than to round down, or vice versa, the observed mean may not be greatly distorted.

The Reported Date of Birth

The proportions still breastfed tabulated by the number of months elapsed since the births in question, do not show such large drops at multiples of 6 and 12 months as one might expect if the heaping in the retrospective data were genuine. But this is not conclusive proof that the heaping is artificial, for the proportions still breastfeeding may themselves be flawed by errors in the reported dates of birth.[18] Many types of error would in fact lead to a smooth-

[17] In the first place, detection of inconsistencies in the data is one-sided: overstatement of breastfeeding durations may lead in some cases to visible inconsistencies (breastfeeding durations in excess of the current age of the child or of its age at death), whereas understatement will not lead to any manifest inconsistency. Secondly, even if an inconsistency is detected, it is often impossible to determine which is most probably in error – the breastfeeding report or the conflicting date of birth or age at death.

[18] In Korea, however, where dates of birth are rather reliably reported, we can note that the presumably reliable current status data indicate very little concentration of weaning at 12, 24, etc months, whereas the retrospectively reported durations have very marked peaks at these points.

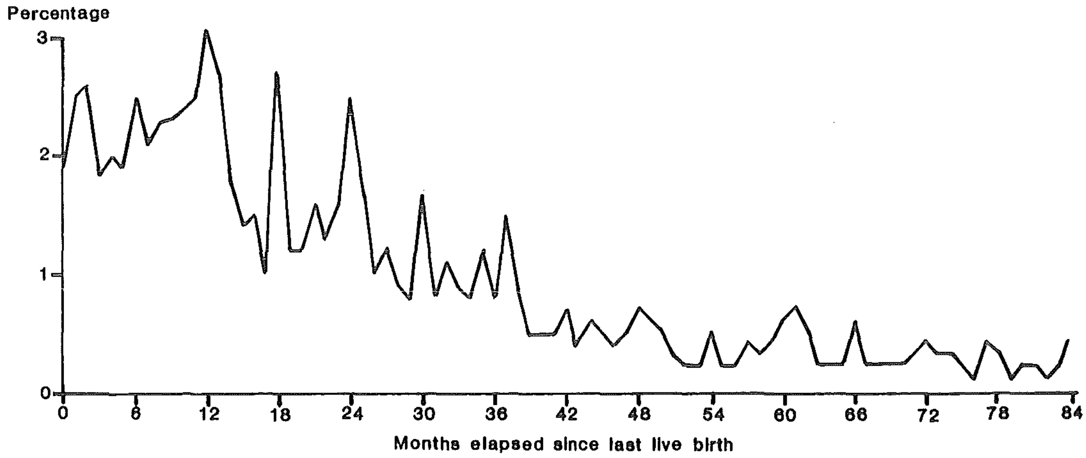


Figure 5 Percentage distribution of women by reported date of their most recent ('last') live birth (PFS)

ing out of any genuine concentrations that might exist. For example, incomplete or missing information on birth dates is common in WFS surveys (Chidambaram *et al* 1980); in Pakistan, 10 per cent of the most recent births (and 20 per cent of all births) had only the year of occurrence recorded in the field and the calendar month was imputed subsequently in the office. Imputation of dates in WFS in general was either made on a strictly random basis or based on a related algorithm, both within the limits imposed by such constraints as the minimum plausible interval between successive births (see Guidelines for Data Processing). Imputation of the month within a given calendar year could thus have shifted individual birth dates from 1 to 12 months in either direction. This would automatically tend to smooth out the apparent patterns of weaning.[19]

The proportion of dates imputed in the office undoubtedly gives a too optimistic picture of the actual amount of approximation embodied in the data in most populations. Many of the birth dates recorded in the field are no doubt arrived at by 'negotiation' between interviewer and respondent as to what date is most plausible, or by plain guesswork. Thus in Pakistan, although 90 per cent of the most recent births had both month and year recorded in the field, the data show clearly that approximations were not uncommon. Despite the fact that the questionnaire posed the question in terms of year and month of birth, it is clear that many of the answers recorded were derived from an approximate estimate of the child's age: the ages implied by subtraction of reported date of birth from date of interview show small but distinct peaks at 1, 1.5, 2 and 2.5 years of age (figure 5).[20] Like imputation, this sort of approximation could distort the *shape* of the distribution although it would not necessarily seriously distort the mean.

More serious would be any systematic tendency for births to be reported as having occurred more recently than in fact they did (or as having occurred in the more distant past than they actually did). Such systematic misstatements would result either from misreporting or from inappropriate imputation algorithms.[21] Indeed, in Pakistan there may well have been a systematic tendency to report or to impute rather recent births as having occurred more recently than in fact they did. Examination of the distribution of births by number of months between the recorded date of birth and the date of interview shows that 13 per cent more births were recorded as having occurred within the year before the survey than were recorded on average for each of the three preceding years. The excess varies slightly between socio-economic groups (14 per cent for women in rural areas compared with 9 per cent for urban areas, 13–14 per cent for births to illiterate women compared with 11 per cent among births to women with primary education) but falls to near zero only for births to women with post-primary education. Since a recent upsurge in fertility of this magnitude is unlikely to have occurred, we conclude that fairly widespread mislocation of dates may have occurred, which could affect our analyses.

Conclusions

Which type of data we would prefer to use if we had the choice would depend on our assessment of which type we think is subject to the least serious errors. For example, we might well put greater faith in analyses drawing heavily on current status data in populations like Korea where dates of birth are relatively reliably reported, or for those duration variables like amenorrhoea that we think are not recalled easily.

In many instances we do not really have a free choice, however. If we want to look at a sample of closed birth intervals or births in the more distant past we have only

[19] For example, if there were a genuine concentration of weaning at 24 months, then the reported proportions reported as still being breastfed among children just under 24 months old would be artificially deflated by inclusion in this group of some children who were in fact older. Similarly, the apparent proportion still being breastfed among those born slightly more than 24 months before the survey would be artificially inflated by the inclusion of some children born more recently.

[20] Beyond 2.5 years this regular peaking pattern disappears, probably because the proportion of non-responses, and hence of 'random' office imputation of calendar month, tends to be more common for events further in the past.

[21] For example, Chidambaram and Pullum (1981) have shown that the estimated time distribution of recent births in Bangladesh, where a large proportion of dates were imputed, depends very heavily on whether the imputation algorithm used assumes that when children's ages are reported in years they are reported in completed years or are rounded to the nearest number of whole years.

Table 1 Selected socio-economic and demographic characteristics by region and place of residence: ever-married women under age 50 (PFS)

Background characteristics	Region												All Pakistan
	Punjab			Sind			NFWP			Baluchistan			
	Rural N = 2586	Urban N = 741	Total N = 3327	Rural N = 625	Urban N = 455	Total N = 1080	Rural N = 397	Urban N = 66	Total N = 463	Rural N = 61	Urban N = 21	Total N = 82	Total N = 4952
Mean age at first marriage	15.7	16.0	15.8	14.6	15.3	14.9	15.1	16.0	15.2	14.1	16.3	14.7	15.51
Mean number of children ever born	4.2	4.4	4.2	3.9	4.3	4.1	4.1	4.2	4.1	3.4	3.7	3.5	4.17
% ever used contraception	6.7	21.2	9.95	2.7	20.9	10.4	7.5	20.6	9.5	*	*	*	9.87
(absolute number in sample)	(174)	(157)	(331)	(17)	(95)	(112)	(30)	(14)	(44)				(489)
% literate	6.2	26.8	10.8	2.3	25.4	12.0	4.2	28.9	7.8	*	*	*	10.68
(absolute number in sample)	(159)	(199)	(358)	(14)	(116)	(130)	(17)	(19)	(36)				(529)
% exposed to mass media	23.4	51.8	29.7	42.1	62.3	50.6	29.8	48.5	32.6	45.1	64.5	51.2	34.87
(absolute number in sample)	(604)	(384)	(988)	(263)	(284)	(547)	(119)	(32)	(151)	(28)	(14)	(42)	(1727)
% ever worked	14.7	22.7	16.5	47.7	23.0	37.3	19.0	14.4	18.4	17.6	35.5	21.9	21.30
(absolute number in sample)	(381)	(168)	(549)	(298)	(105)	(403)	(75)	(10)	(85)	(11)	(7)	(18)	(1055)

*Number of cases in numerator 5 or less.
NOTES: The figures for absolute frequencies are rounded to the nearest whole number.

retrospectively reported durations. If we choose to look at very recent births, some of the women will still be breastfeeding and cannot give a reported complete duration. In practice, the choice of data type is often subordinate to the choice of which births to analyse, a decision with important implications in its own right that is discussed in chapter 4. Awareness of the typical potential weak spots in the different data types outlined here is, however, essential if we are to avoid over-interpretation of whatever data set we use.

3.4 THE PAKISTAN FERTILITY SURVEY

The Pakistan Fertility Survey (PFS) was conducted in 1975. The questionnaire was essentially the WFS core questionnaire adapted to Pakistan conditions. Since the module on factors other than contraception affecting fertility was not used, the data for Pakistan – like those for many countries – include no information on full breastfeeding nor specific questions on post-partum amenorrhoea or post-partum abstinence. The Pakistan data set is, however, more comprehensive with respect to breastfeeding than the data obtained from the core questionnaire in most countries. Instead of asking the two basic questions about breastfeeding just for, at most, two births per woman, the PFS asked them for *all* births, the questions being included as part of the full maternity history asked of every woman interviewed. This additional information allows us to escape from some of the constraints imposed when data sets are restricted to just the last one, or two, births for each woman. Although it is questionable whether women can report breastfeeding durations retrospectively with sufficient accuracy for us to detect trends unless those trends are dramatic, there are several other ways we can use the additional data. In particular we shall use them here to illustrate more complete methods of estimating recent breastfeeding patterns and to illustrate some of the consequences of using data sets that are restricted to just the last one, or two, births per woman.

A short description of the sample is necessary before we proceed to the actual analysis. More detailed information can be found in the PFS First Report (Population Planning Council of Pakistan, 1976).

The sample represents about 92 per cent of the total population of Pakistan. A number of areas inhabited by unsettled nomadic and tribal populations were excluded from the original sample design, as were areas that were sparsely populated and highly inaccessible;[22] these areas account for nearly 7 per cent of the national population. We can note that, because of these exclusions, the data are concentrated on the more settled population and exclude several of the smaller ethnic groups with different life styles. The remaining 1–1.5 per cent of the population that was not covered by the survey live in parts of Baluchistan Province where it was not possible to conduct the interviews planned because of inaccessibility.[23] Their exclusion probably has rather little effect on national estimates, but since Baluchistan Province constitutes less than 4 per cent of the national population at any rate, exclusion of 1–1.5 per cent makes the achieved sample for this province very small (less than 100 women) and not necessarily representative of the province. For some of the detailed analyses we have, therefore, excluded certain subgroups, notably Baluchi and Barochi-speaking women.

A two-stage[24] stratified sampling procedure was used to select ever-married women under age 50. Disproportionate sampling was used, with a higher sampling fraction in urban than in rural areas. The data used in this report are the data from the Preliminary Standard Recode Tape, and the analyses incorporate appropriate weights to compensate for the disproportionate selection, so that rural and urban areas are appropriately represented. The effective sample size totals 4952 women.

Table 1 presents a summary of half a dozen major socio-economic indicators for these women. The sample is rather homogeneous – overwhelmingly Muslim, predominantly illiterate (89 per cent) and largely rural (74 per cent). Their average age at first marriage was low (15.5 years), their fertility quite high (an average of 4.2 children to date for women of all ages). Few major regional differences are apparent, except for the indicators of exposure to mass media and work (for both of which Sind Province scores markedly above the average). Rural-urban differentials are strong for those two indicators and also for literacy and for use of contraception, but they are not strong for either age at marriage or cumulated fertility.[25]

[22] Excluded were restricted cantonment areas and some former states and tribal areas of the NWFP-Swat, Dir, Chitral, Malakand Agency, Kurram Agency and Khyber Agency.

[23] The study population therefore excludes in addition rural areas of Kalat, Mekran, Loralai, Zhob and Kharan Districts of Baluchistan Province.

[24] In Baluchistan, three-stage.

[25] This table summarizes the characteristics of women for whom detailed information was collected – that is , in the PFS, ever-married women. Average age at marriage (and also fertility) exhibit larger urban-rural differentials when all women – including women who are not yet married – are considered.

4 The Estimation of Breastfeeding Patterns

4.1 PRELIMINARY CONSIDERATIONS AND THE CHOICE OF AN ANALYTICAL STRATEGY

Before deciding on a particular method of estimation we have to answer two questions:

1 What exactly is it about breastfeeding that we want to measure?
2 For which births do we want to measure it?

The Choice of Parameters to be Estimated

Our answer to the first question usually falls simply and squarely on the frequency distribution for the duration of breastfeeding. We may be interested in estimating the entire distribution or we may be interested in estimating only its summary statistics, primarily its central tendency, perhaps also a measure of its dispersion. In some cases we may be interested in addition in a particular section of the distribution – for example, the proportion who did not breastfeed at all.

Only if we restrict ourselves to births where all the children have been weaned before the survey (for example, to data from the last closed birth interval for each woman), do we have a report on the total duration of breastfeeding for each child concerned and hence the material for constructing a frequency distribution directly. Unfortunately, restricting ourselves to births where weaning has already occurred introduces selection biases that can reach serious proportions. We should, therefore, use other birth sets in most cases, in which some of the children concerned have not yet been weaned. We know that these children have been breastfed a certain number of months already, but we do not know how much longer they will be breastfed. It is, therefore, necessary to use estimation methods appropriate for censored data. The various methods available differ in their data requirements (as well as differing slightly in their assumptions), and so the choice between them will depend in part on the availability and quality of different types of data. In addition, however, none of these methods provides a direct estimate of the frequency distribution. The frequency distribution or its characteristics must be obtained indirectly, and the different methods we can use vary considerably in their capacity to yield details of the distribution. If we are interested only in the mean, then we may be able to use very simple estimation procedures. If the median is required or measures of dispersion, such as the standard deviation or the interquartile range, then we must embark on more complex methods.

For Which Births Do We Want to Estimate Breastfeeding Experience?

Table 2 summarizes the pertinent characteristics of the data sets available in WFS surveys in general. For most countries we have breastfeeding data referring to two main types of birth sets:

Type I: the two most recent births per woman
I(a): the last closed birth interval for each woman
I(b): the current open birth interval for each woman
Type II: all intervals (whether open or closed) following births that occurred in a given period immediately preceding the survey.

These different birth sets can exhibit widely divergent breastfeeding patterns. They are, in fact, conceptually quite distinct from each other. The first group (and particularly the last closed birth interval) was especially stressed in both the design of the questionnaire and the early data processing. In this section we shall outline why the second set is preferable for many purposes.

Choice of a General Strategy

An underlying conceptual issue in our choice of strategy is whether we are more interested in examining breastfeeding patterns from the *women's* viewpoint or from the *children's.* In the former case, we would want to examine the distribution of women by their breastfeeding behaviour; in the latter, the distribution of children by how long they are breastfed. Which approach is more appropriate may depend on how we want to use the information. For example, if we are concerned about the impact of breastfeeding on infant survival chances, then the children may seem the more natural unit of analysis, whereas if our concern lies in disentangling the social or other determinants of women's breastfeeding behaviour, then women may seem to be the more natural unit of analysis.

In general we might expect the average birth interval per woman to be longer than the average birth interval per child, since women with short birth intervals contribute more children than women with long birth intervals. If – as is usually the case – duration of breastfeeding is positively correlated with length of the birth interval, then we would expect average breastfeeding durations measured per woman to be longer than the averages per child.

In practice, however, the distinction is usually blurred and our choice made more complicated by the fact that the available data refer to a restricted subsample of births and are not necessarily representative either of all women or of all children. Selection biases can then make unnuanced interpretation of the results quite hazardous.

Analysis of Data for Women, Using Data for the Two Most Recent Births per Woman

We have already referred to some of the general limitations of data sets restricted to the two most recent births, par-

Table 2 Characteristics of breastfeeding data sets in WFS surveys: the sample of births to which each refers and the type of data

Birth set	No fixed time period			Data referring to a specified time period	
	I Two most recent births for each woman			II All births in the Z years[a] immediately preceding the survey	
	I(a)	I(b)	I(c)	II(a)	II(b)
Birth set					
Births included	Next to most recent birth ('last' closed birth interval)	Most recent birth (current open birth interval)	Two most recent births weighted inversely according to whether women reported on 1 or on 2 births	All births in period, weighed equally	All births in period, weighted inversely according to no of births woman had in the period[b]
Sample to which the data refer	Breastfeeding behaviour of women who had had at least 2 children before the survey	Most recent breastfeeding behaviour of women who had had at least 1 child before the survey		Breastfeeding experience of children born in the period	Breastfeeding behaviour of those women who gave birth in the period
Type of data					
Duration of breastfeeding (reported retrospectively)	Available for all the births[c]	Available only for births where child was weaned before the survey		Available only for births where child was weaned before the survey	
Current breastfeeding status and date of birth	Non-informative[c] (children already weaned)	Available for all the births[d]		Available for all the births[d]	

[a]Z is most conveniently defined as equal to or longer than the longest duration of breastfeeding in the population.
[b]An alternative is to select at random one birth that occurred in the period for each woman.
[c]We assume here that women do not continue breastfeeding a child through their next pregnancy.
[d]Available either directly or by inference (assuming that only the most recently born child can still be breastfeeding).

ticularly to the point that they cannot yield a sample that is representative of any given time period, except for a subsample representing a very short (often too short) period immediately preceding the survey.

We should mention at least two other points here, both related to problems in using either of these intervals for a per woman analysis. Both are very well known in birth interval analysis and have already received detailed development elsewhere,[26] but both have tended until recently to get pushed into the background in empirical analyses of post-partum variables.

Data for the last closed birth interval under-represent women with longer than average birth intervals because these women have had less chance of having reached their second birth. In most developing countries duration of breastfeeding is positively associated with length of the birth interval, so this selection usually translates into an under-representation of women with long durations of breastfeeding.

Data for the current open birth interval avoid this particular trap (provided of course that we do not exclude those women who are still breastfeeding), but raise other potential problems of their own. For example, a woman does not have exactly the same interval between each pair of successsive births. The probability that a single-round survey will be held during a particular interval is proportional to the length of that interval. Thus those intervals that are longer than a woman's average have a higher chance of being cut by such a survey and hence of being recorded as the current open birth interval than do her shorter birth intervals.[27, 28]

To sum up, neither the current open birth interval nor the last closed birth interval yields unbiased samples of birth intervals per woman and hence neither yields unbiased samples of breastfeeding durations wherever breastfeeding is correlated with birth interval length. Figure 6 illustrates how different the results can be depending on which birth set one uses. The two solid lines give the estimated proportion of women still breastfeeding d months after the birth derived from the data for the last closed interval and from the current open interval (standard definitions). The estimated mean duration of breastfeeding is over 6 months shorter in the former than in the latter! It is even 3–4 months shorter than the estimate from the proportions of women still breastfeeding derived from type II data (the

[26]Notably in the work of Louis Henry (Henry 1972, for example) and of Mindel Sheps and Jane Menken (1973).
[27]For an intuitive discussion and illustration of other potential problems associated with analysis of breastfeeding data for the current birth interval, see Page *et al* (1980).
[28]To the extent that intra-woman variation in intervals and any related intra-woman variation in breastfeeding depend on the survival of the first of the two children, we can reduce this effect by restricting analysis to intervals where the child survived.

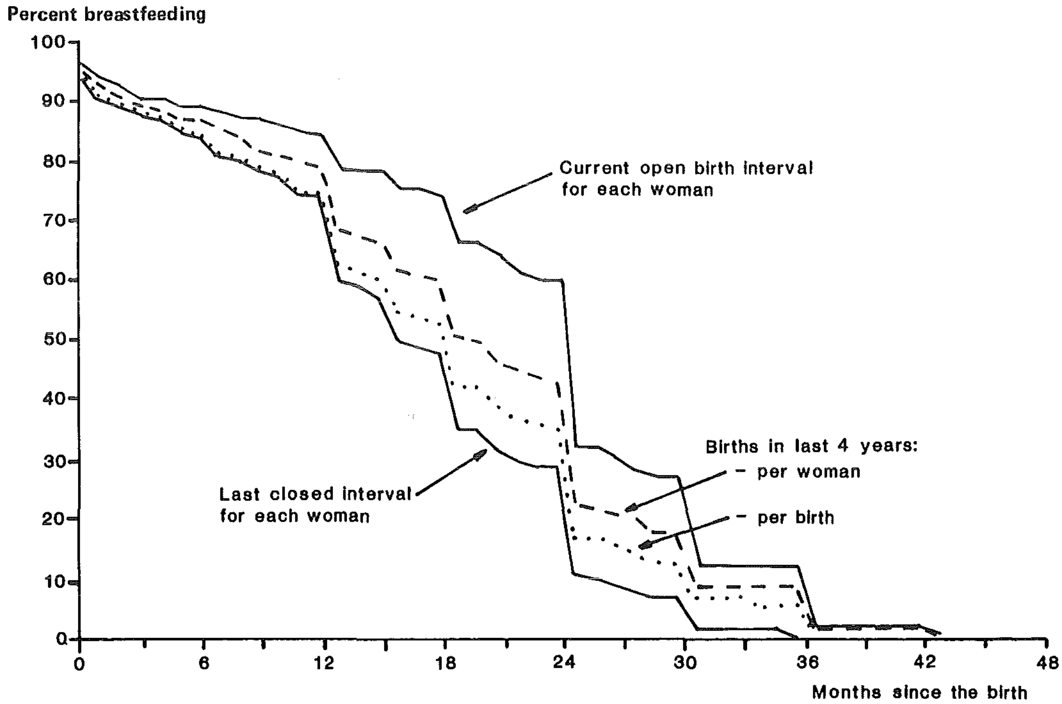


Figure 6 Comparison of estimated proportions breastfeeding, by the number of months elapsed since the birth, derived from four different birth sets (PFS)

broken line). Since the two data sets are biased in opposing directions, one could reduce the overall bias by combining them (type I(c) in table 2), but the separate biases are unlikely to cancel each other out completely (and we would not always even be sure of the direction of the net bias). Perhaps even more seriously we would still have a data set that does not refer to the same time period for all women. In populations where breastfeeding may have been changing over time, this lack of a fixed time reference may make use of these data highly questionable.

Analysis of Data for All Births in a Given Period

In contrast, data for all births in a given period immediately preceding the survey (type II) are relatively clean. We must, however, keep in mind that they refer to a sample of children not to a sample of women and must be interpreted accordingly. For example, if we find breastfeeding differentials associated with characteristics XYZ of the mother, these differentials do not measure the difference in breastfeeding behaviour of women with characteristics XYZ but the difference in breastfeeding experience of children born in the given period to women with these characteristics.

If an analysis for women rather than for births is required, then we can take appropriate steps. The fact that the birth set includes more births for women with short birth intervals than for women with long ones can be allowed for by weighting each birth in inverse proportion to the number of births experienced by that child's mother during the period concerned (the broken line in figure 6 was derived in this way); alternatively, for those women who had more than one birth during the period we can select just one birth at random. Figure 6 also illustrates the difference between the two approaches. Since breastfeeding is positively correlated with birth interval length in Pakistan, giving each woman rather than each birth equal weight produces higher proportions still breastfeeding at each duration (d) since the birth and a higher mean duration of breastfeeding. Note that the results are restricted to those women who had at least one birth during the period concerned, rather than representing all women. Among older women, only those who have continued childbearing to relatively late ages are included and, more generally, in any group women with particularly long birth intervals will be under-represented (this will be especially obvious if the period covered is short). We do, however, retain all the advantages of a fixed reference period.

Conclusions

Clearly, analysis of the breastfeeding data derived from all births in a given period immediately preceding the survey is the preferred strategy for most purposes. The analyses that follow adopt this strategy. Of the two possible approaches based on these data (giving each birth equal weight versus giving each woman equal weight), we personally prefer in most instances to use the former.

4.2 ESTIMATION OF BREASTFEEDING PATTERNS IN A PERIOD IMMEDIATELY PRECEDING THE SURVEY

In this section we illustrate ways of estimating breastfeeding durations for a period of Z years preceding the survey, where Z is the longest duration of breastfeeding in the population. The next four sections illustrate the estimation of a breastfeeding life table or of the frequency distribution of breastfeeding durations. A much simpler procedure that can be used if only an estimate of the mean duration of breastfeeding is required, is illustrated below on pp31-3.

Introduction: Analysis of Censored Data on Breastfeeding

The classic procedures for handling censored data for a non-renewable event like cessation of breastfeeding are the well-known techniques of life-table construction that take into account the duration of exposure of the censored cases.[29] Where the duration of exposure (here, the number of months elapsed since the birth in question) is known for all cases and the time to event or 'survival' time known for all those who have already experienced the event (here age at weaning for those children who have already been weaned), one can easily estimate the conditional probability of experiencing the event (weaning) between exact durations x and (x + 1) as

$$ {}_1q_x = {}_1D_x / {}_1N_x \qquad (1)$$

where ${}_1N_x$ is the number of cases of who have been exposed throughout the interval x to (x + 1) to the risk of experiencing the event and ${}_1D_x$ is the subset of these cases who actually experienced the event during this interval.

From the ${}_1q_x$ values, the estimated probability of not experiencing the event before exact age x, the 'survivor' function, can immediately be calculated as

$$ l_x = \prod_{i=0}^{x-1} (1 - {}_1q_i) \qquad (2)$$

The first differences yield the estimated proportions, ${}_1d_x$, experiencing the event in the interval between x and (x + 1). Assuming that on average they experience the event at the mid-point of the interval, the mean time elapsed before experiencing the event is estimated as

$$ e_0 = \sum_{x=0} {}_1d_x (x + 0.5) \qquad (3)$$

Alternatively the mean duration can be estimated directly from the l_x values, making the same assumption, using the equivalent expression

$$ e_0 = 0.5 + \sum_{x=1} l_x \qquad (4)$$

The analysis of breastfeeding differs, however, from classic life-table methods in one important respect. The construction of decrement tables conventionally applies to a real or synthetic cohort of persons, all of whom are exposed to the risk of experiencing the event in question; the above expressions exclude any cases that were never exposed to this risk. For mortality, for example, life tables refer conventionally to live births: still-born children (children with a duration of life of exactly zero) are excluded. For a number of phenomena, however, of which breastfeeding is one, the subgroup with exact duration zero is also of interest. This category should not be excluded when we are calculating the overall mean duration of breastfeeding, for example. We are often more interested in estimating the distribution by duration of breastfeeding (including those who were never breastfed) than in the distribution by age at weaning among those at risk of weaning (ie restricted to those who were breastfed). Expressed another way, instead of being most interested in the proportions *still* being breastfed at age x months among those children who were breastfed, we are more interested in the proportions being breastfed at age x months among all the children in question, whether or not they were breastfed. We shall refer to this latter proportion as l'_x. It is defined simply as

$$ l'_x = E \prod_{i=0}^{x-1} (1 - {}_1q_i) \qquad (5)$$

where E is the proportion ever breastfed

The first differences, ${}_1d'_x$, give the proportions breastfed for each interval x to (x + 1), whom we assume to have had an average breastfeeding duration of (x + 0.5) months. (1 – E) gives the proportion with a duration of breastfeeding of zero. The mean duration of breastfeeding can, therefore, be estimated as

$$ \bar{Y} = (1 - E).0 + \sum_{x=0} {}_1d'_x (x + 0.5) \qquad (6)$$

Alternatively the mean duration can be calculated directly from the l'_x values as

$$ \bar{Y} = 0.5\,E + \sum_{x=1} l'_x \qquad (7)$$

$$ \text{or } E\,[0.5 + \sum_{x=1} l_x] $$

Direct application of these procedures is illustrated in the next section. For many countries, however, direct applications are not possible either because data are not available for all the births or because the retrospectively reported durations of breastfeeding are not considered to be sufficiently reliable. In these circumstances, the current status data should be used instead, and the reader interested only in that can skip immediately to 'Estimation of a Breastfeeding Life Table from Current Status Data', on p25.

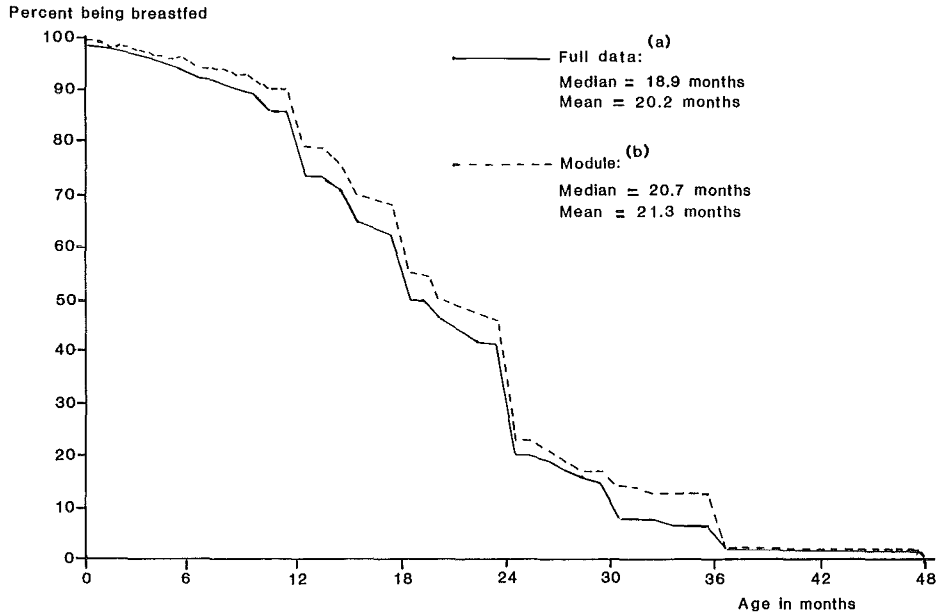
Estimation of a Breastfeeding Life Table when Duration of Breastfeeding Was Asked for All the Births

When data are available for *all* children born in a suitable period immediately preceding the survey (age at weaning for those already weaned, duration of exposure for those still being breastfed), it is a straightforward, if somewhat laborious, task to estimate the conditional probabilities of weaning, ${}_1q_x$, and hence a full breastfeeding life table. Note that duration of breastfeeding has to have been asked about all births in at least the Z years preceding the survey.[30] In Pakistan, since breastfeeding questions were asked about every child each woman had borne, we can use this approach.

Strictly speaking the data should be in the form of completed months of breastfeeding and completed months of exposure, which is not the case with WFS data sets. It is obvious that the retrospectively reported durations of breastfeeding were given in rounded numbers in most cases.

[29] For a general discussion of the application of life table analysis to a range of duration variables in WFS data sets, see Smith 1980; for a detailed application to birth interval analysis, see Rodríguez and Hobcraft 1980.

[30] The analysis can either be made just for the Z years or (if the data are available) for a longer period if the sample sizes would otherwise be too small and if breastfeeding patterns have not been changing.



NOTES:
(a) Estimates obtained using data for all births in the period.
(b) Estimates that would have been obtained if data had been available only for the two most recent births for non-pregnant women and for the one most recent birth for pregnant women.

Figure 7 Proportions being breastfed by age, derived from estimated conditional probabilities of weaning: surviving children born in the four years preceding the survey (PFS)

The durations of exposure, measured as the difference between calendar month of interview and calendar month of birth, do not measure durations in completed months either. Instead they yield categories of overlapping cohorts (a difference of 1 calendar month could be a child born 1 day before the survey (0 completed months) or 60 days before (1 completed month)) centred on the duration assigned. On average the durations assigned are approximately centred on exact durations rather than being measured in completed months. The estimates obtained from such data refer approximately to ages roughly one half a month less than the conventional ages. In other words, we obtain approximate estimates of $l_{x-0.5}$ rather than l_x (Smith 1980).

Figure 7 shows the results for all Pakistan. Most noticeable is the stepped pattern reflecting the extremely strong heaping of the retrospectively reported breastfeeding durations on multiples of six months, a significant proportion of which is probably an artefact resulting from digit preference in the responses. Not noticeable would be any systematic tendency towards over- or understatement of breastfeeding durations or of the durations of exposure. We have already mentioned the fact that the reported dates of birth suggest that too large a proportion of recent births may have been reported as having occurred during the last year. If misreporting were more common for children being breastfed than for weaned children, or if the reported duration of breastfeeding for weaned children was understated to correspond with any understatement of their age, this could lead to a poor estimate of the $_1q_x$ values and to a poor estimate of the l_x values and of the mean duration of breastfeeding.

This method is not strictly applicable for most WFS data sets (Rodríguez and Hobcraft 1980) because information is lacking for some of the births that occurred during the period concerned: the subset of births for which breastfeeding questions *were* asked is biased towards longer than average birth intervals. To investigate the impact of this selection, we have taken the Pakistan Fertility Survey data, and simulated what would have happened if only the two most recent births per woman had been included (as in the standard core questionnaire) or if this pattern had been followed for non-pregnant women but only one birth had been included for pregnant women (as in the FOTCAF module). For Pakistan at least, the resulting errors in the breastfeeding estimates would not have been very large (figure 7). Use of the FOTCAF module, for example, would have led to a 1 per cent overestimate in the proportion of children breastfed (95.1 per cent instead of 94.3 per cent), and a 1–2 month overestimate of the average duration of breastfeeding. To obtain an unbiased data set we must use the current status data.

Estimation of a Breastfeeding Life Table from Current Status Data

For most WFS data sets the only breastfeeding data available for all the births in the Z years preceding the survey are current status data. Fortunately their use is relatively straightforward. Moreover, they are not affected by digit preference in the same way as retrospectively reported durations for the post-partum variables.

We have already mentioned that the difference between calendar month of interview and calendar month of birth yields single month categories of duration of exposure (each formed of two birth cohorts) centred on the exact duration corresponding to the difference between the two calendar months. In other words, we can usually assume that those recorded as born d months before the survey

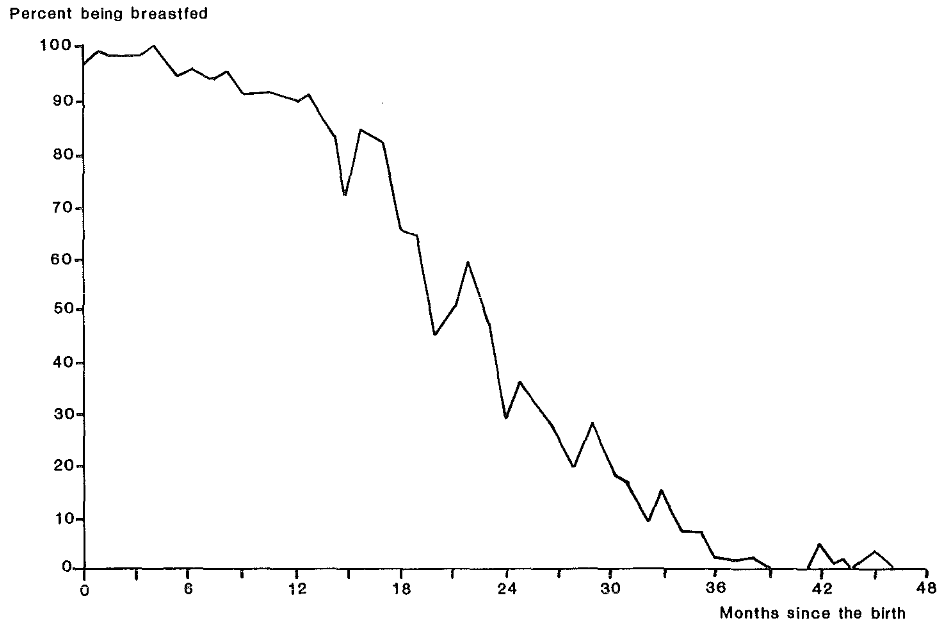


Figure 8 Current status data – proportions being breastfed by the number of months elapsed since the birth: surviving children born in the four years preceding the survey (PFS)

were, on average, born exactly d months ago.[31] The proportion currently being breastfed among those recorded as born (d) months ago, P(d), can be taken as an estimate of l'_x (x = d) for that cohort of children.[32] We can, therefore, use the reported proportions P(d) to estimate life-table functions. Note that this holds only for $d \geqslant 1$: $d = 0$ in this convention refers to children born in the calendar month of interview, who are a subset of the children between exact ages 0 and 1, not to children at exact age zero.

The P(d) values can be taken directly as estimates of l'_x. However, unlike the l'_x estimates obtained via estimates of the conditional probabilities, $_1q_x$, the series of estimates thus obtained does not necessarily form a monotonically declining sequence. In practice it is usually extremely irregular as figure 8 illustrates. The reason is that the sample is heavily fragmented: each P(d) value is based on just the small subsample of births that occurred d months ago. Even at national level in Pakistan with a total sample size of nearly 5000 ever-married women, there were only about 100 births reported on average per month, from which to calculate single month values of P(d). Some months had far fewer than 100, largely because of rounding in the reported ages of the children on multiples of six months, which tends to evacuate non-preferred values of (d). The result of this sample fragmentation is considerable sampling variability and an irregular P(d) sequence even at national level; when we look at subgroups within the population, the irregularity can become even worse.

If the P(d) series does not decline monotonically, we cannot take the first differences to estimate $_1d_x$ values and hence a standard deviation. Nor can we always identify a median (or other quartiles) because the sequence may pass the 0.50 (0.25, 0.75) level at more than one point (in figure 8, for example, there is no unique estimate of the median because the sequence crosses 0.50 not once but three times). It is, however, still possible to calculate a mean directly from our raw P(d) values despite irregularities, using equation (7) which here becomes

$$\bar{Y} = 0.5\,E + \sum_{d=1} P(d) \qquad (8)$$

provided there are observations for all values of d.[33] If any cell is empty (a month for which no birth at all was reported for the population or subgroup in question), then we cannot calculate the mean directly from the raw data using equation (8).[34] In such circumstances we must either group the single month categories into broader groups to eliminate the empty cells, or use some form of smoothing. Even if none of the cells are empty, some form of smoothing may be desired to reduce irregularities believed to be merely the result of sampling variability before computing an estimate of the mean.

The simplest procedure is to block the data into broader categories for (d) – for example categories 3 months wide centred on 3, 6, 9, etc months. If we use categories of width n months, centred on d, then the proportion currently breastfeeding is calculated (for n taking odd number values) as:

$$P'(d) = \frac{B\left(d - \frac{n-1}{2}\right) + \ldots + B(d) + \ldots + B\left(d + \frac{n-1}{2}\right)}{N\left(d - \frac{n-1}{2}\right) + \ldots + N(d) + \ldots + N\left(d + \frac{n-1}{2}\right)} \qquad (9)$$

Three-month wide blocks are often wide enough, and we

[31] Assuming a constant stream of births and of interviews.

[32] If births and interviews are evenly spread and if the true proportion being breastfed declines linearly with age around x, then it is an unbiased estimate.

[33] If the questions on breastfeeding duration were not asked for all the births in the period concerned, then E must be estimated from all births in a shorter period for which information *is* complete.

[34] A common mistake is to assign a value of zero to empty cells, when in fact the P(d) value must be treated as unknown (a 'missing value') not as a zero value.

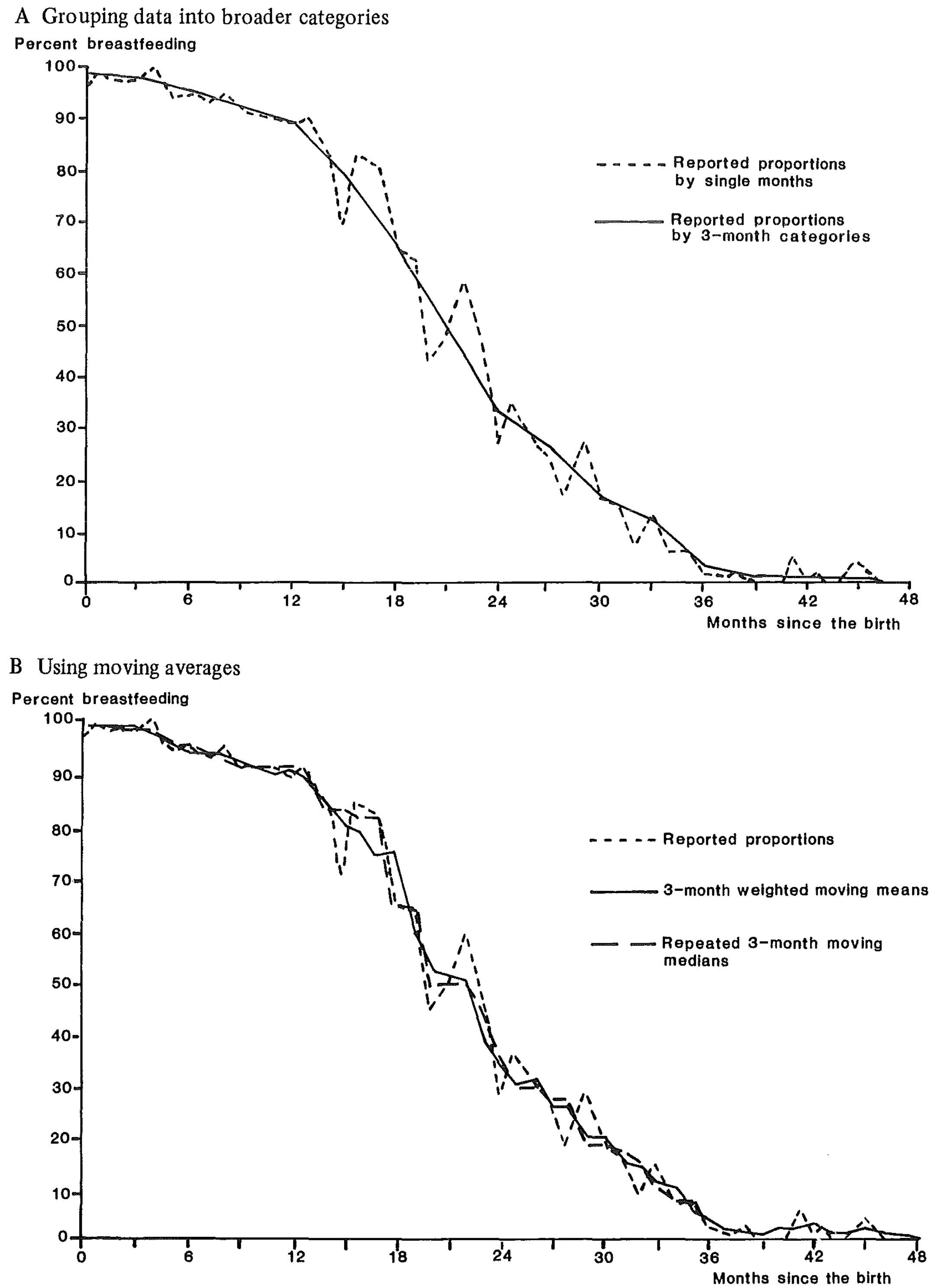


Figure 9 Smoothing current status data

have

$$P'(d) = \frac{B(d-1) + B(d) + B(d+1)}{N(d-1) + N(d) + N(d+1)} \qquad (10)$$

The mean duration of breastfeeding is then estimated, assuming a linear decline of P(d) within each block, as:

$$\bar{Y} = 0.5n.E + n\ \Sigma\ P'(d) \qquad (11)$$

The disadvantages of this method become increasingly apparent as the block sizes become larger. Although the P'(d) values are more stable than the original P(d) values, they are less numerous: the number of points we can estimate on the survivor function is reduced, making it harder and harder to tell the shape of the distribution even if we can estimate its mean. Moreover, we cannot make the blocks too wide, otherwise the assumption of linearity within the blocks may become untenable.

Moving averages provide a means of smoothing (and interpolating) that avoids reducing the number of points for which P(d) can be estimated. We can use either moving means or moving medians.[35] In either case, it may be preferable to use weighted rather than unweighted averages (weighting each of the raw P(d) values by the number of observations on which it is based) if there is strong digit preference in the (d) values. If we do not weight, then the values for a duration such as 24 months, which is typically based on a large number of observations, is given the same weight as the value for unpreferred durations, such as 23 months, which may be based on only a handful of observations. In other words, when the (d) value itself can be in error, it is often better to give each observation equal weight by using weighted moving averages than to give each category of (d) equal weight regardless of the number of observations on which it is based.

The end results of these three smoothing procedures are shown in figure 9. The advantages and disadvantages of each method are clear. Blocking usually produces a generally monotonically declining sequence (especially if block width is large), but a monotonic decline is not guaranteed and the possibility of estimating the more detailed shape of the distribution is lost. Three-month moving means eliminate isolated peaks and troughs (which are almost certainly due to sampling variability) and also round off the corners of any 'steps' in the sequence (which may be due mainly to sampling variability but which could also, conceivably, reflect genuine sharp drop-offs at certain points); the end result is rarely a monotonically declining sequence, tending to show instead several small reversals. Repeated application of three-month moving medians usually comes closer to yielding a monotonically declining sequence (although even here monotonicity is not guaranteed) but leaves any series of irregular sharp drop-offs unsmoothed. If one of these forms of smoothing is to be used the choice between them will depend on

(i) the level of detail required in the final estimates,
(ii) the extent to which a generally monotonically declining sequence is desired, and
(iii) the extent to which other irregularities are thought to be due to noise rather than reflecting real patterns.

None of these methods guarantee a monotonically declining sequence, however, since they all smooth the sequence rather than constrain it. If a monotonically declining sequence is required (to estimate the standard deviation or to obtain unique estimates of particular quantiles) then other procedures must be used that explicitly constrain the sequence. One of the simplest is the 'pool adjacent violators' algorithm (Barlow *et al* 1972). For any sequence of (d) values for which the P(d) estimates rise rather than fall, a single value of P(d) is calculated by pooling the numerators and the denominators (as in equation 9), and this value is assigned to all the cells concerned. As figure 10 shows, this replaces any rising segment by a shelf, leaving the rest of the sequence untouched. Where the raw data are very irregular, with numerous reversals, the end result is a series of little steps.

[35] See Tukey (1977) for a detailed explanation of the use of moving medians. It consists essentially of assigning to each cell the *median* value observed in the group of cells being averaged, rather than the mean value for the cells concerned. The process can be repeated until the sequence stabilizes.

Finally, a more radical solution that not only yields a monotonically declining sequence but also a smooth one can be found using model schedules. A simple relational model schedule of breastfeeding has been developed (Lesthaeghe and Page 1980) such that

$$\hat{\text{logit}}\ P(d) = \alpha + \beta \text{ logit } P_s(d) \qquad (12)$$

where $\hat{\text{logit}}$ P(d) is the estimated logit of the proportion currently breastfeeding at exact duration (d) and logit $P_s(d)$ is the logit of the proportion in a standard schedule.

All we have to do is to plot the logit of the observed P(d) values against logit $P_s(d)$ and fit a straight line to estimate the best-fitting α and β. Given estimates of α and β we can quickly estimate the values of P(d) in the model schedule that best fits our observations:

Since $$\text{logit } P(d) = \ln \frac{P(d)}{1 - P(d)}$$

letting $\hat{Y}$ be the value of logit P(d) implied by our estimates of α and β.

then $$\hat{Y} = \ln \frac{\hat{P}(d)}{1 - \hat{P}(d)}$$

or $$\hat{P}(d) = \frac{e^{\hat{Y}}}{1 + e^{\hat{Y}}} \qquad (13)$$

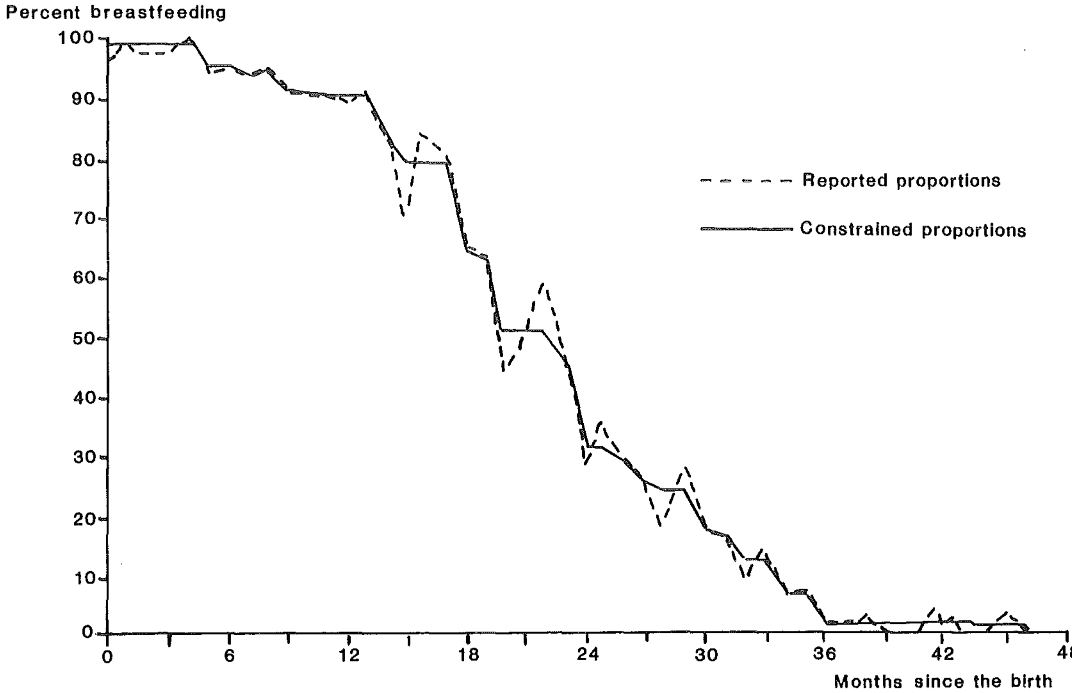
The data for Pakistan linearize well on the standard (figure 11) and the resulting smoothed, declining curve obtained from the estimated α and β is shown in figure 10. This method is particularly useful where it is believed that almost all the irregularities are artefacts rather than genuine, since it smooths out virtually all irregularities. It is also particularly convenient in that the entire schedule can be characterized by just two parameters, α and β.

In general, the various methods of handling irregularities in the current status data yield broadly similar estimates of the mean duration of breastfeeding, although the detailed shape of the estimated distribution can vary. We can thus be fairly confident about our estimate of the average duration, but less so about the overall distribution.

Concluding Comments on Estimation of a Breastfeeding Life Table

The most obvious differences between the two basic methods we have illustrated stem from the effect of heaping in retrospectively reported breastfeeding durations in the first method and irregularities resulting from sample fragmentation in the second. The two methods also differ slightly, however, in the way they would be affected by any change in patterns over time or by misreporting of durations of exposure. This is because the different apparent birth cohorts contribute in different proportions to the final survivor function in the two approaches. This is brought out in figure 12. In the first method, all birth cohorts (except those born in the calendar month of interview) contribute to the estimate of ${}_1q_0$ and hence to l_1 (and, through the cumulative nature of the calculations to subsequent l_x values). The estimate of ${}_1q_1$ excludes the most recent cohort included in the estimate of ${}_1q_0$, and so forth: as x reaches high values only the earliest cohorts and their most recent experience contribute to the ${}_1q_x$ estimate.

A Constraining using 'pool adjacent violators' algorithm



B Constraining and smoothing by identification of best-fitting model schedule

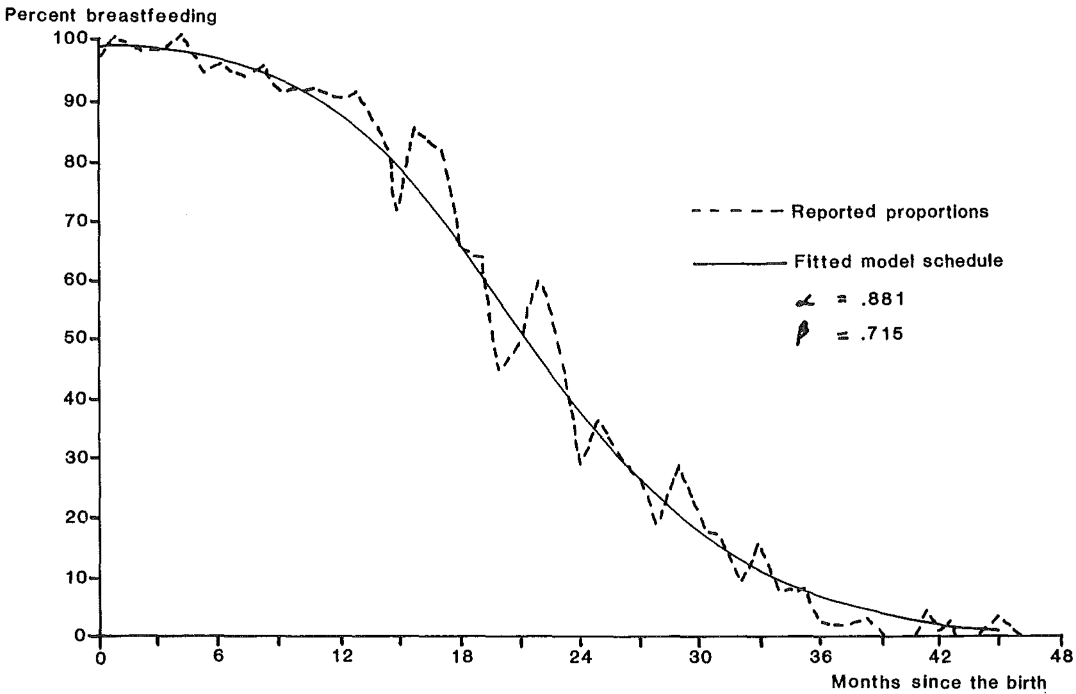


Figure 10 Constraining current status data to produce a monotonically declining sequence of proportions still breastfeeding

By contrast, with the current status data, only recent cohorts contribute to the l'_x estimates when x is low, and as x increases, the l'_x estimates are based on earlier and earlier cohorts and cumulative experience over a longer and longer time period.

All these differences combine to explain the slightly different estimates the two methods yield even of the mean duration of breastfeeding.

Estimation of Only the Mean Duration of Breastfeeding: A Short Method

If we are interested only in estimating the mean and if it is reasonable to assume that the number of births per month has been constant throughout the Z years, then an extremely simple and rather robust estimation procedure can be used (Mosley *et al* 1982).

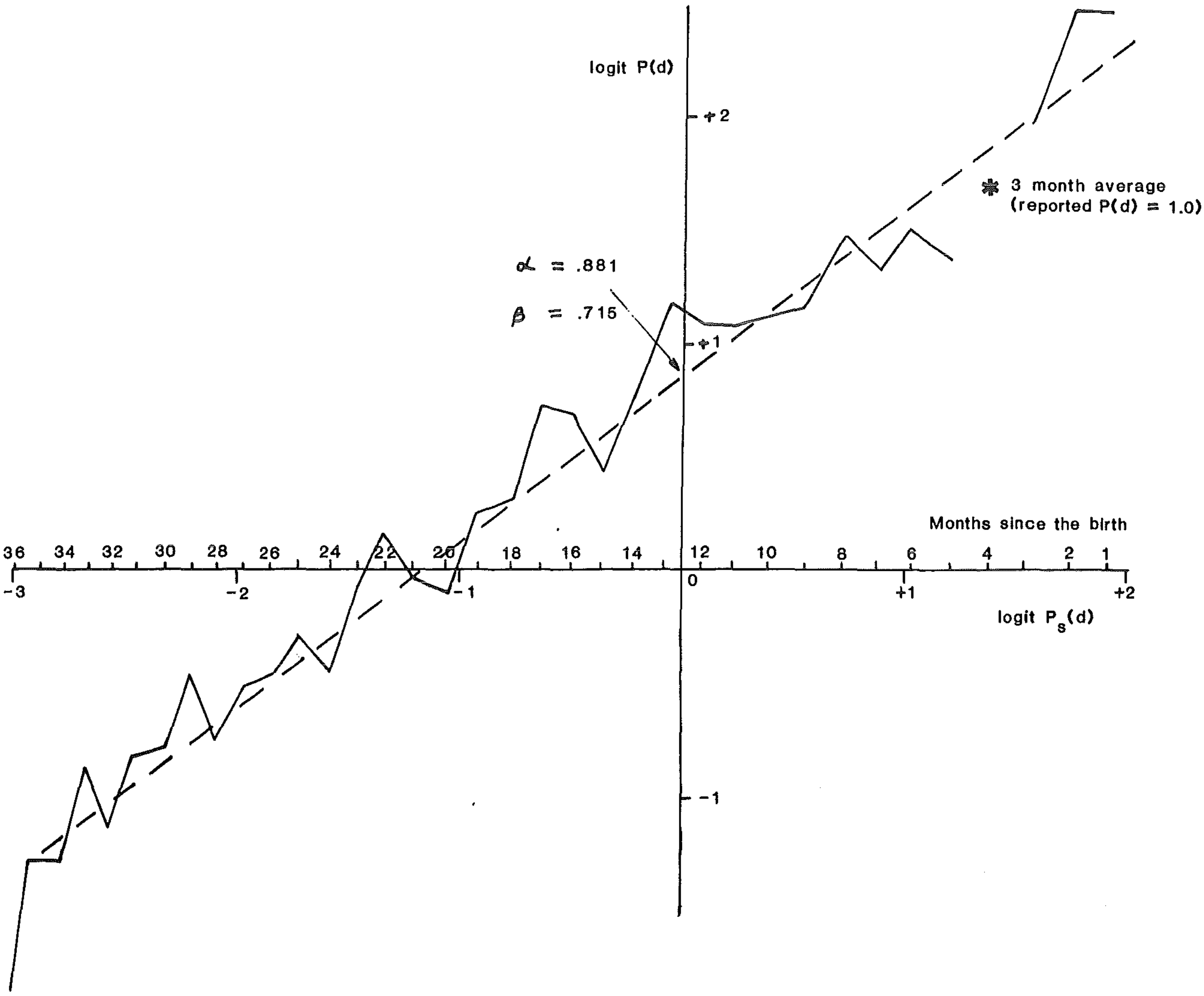
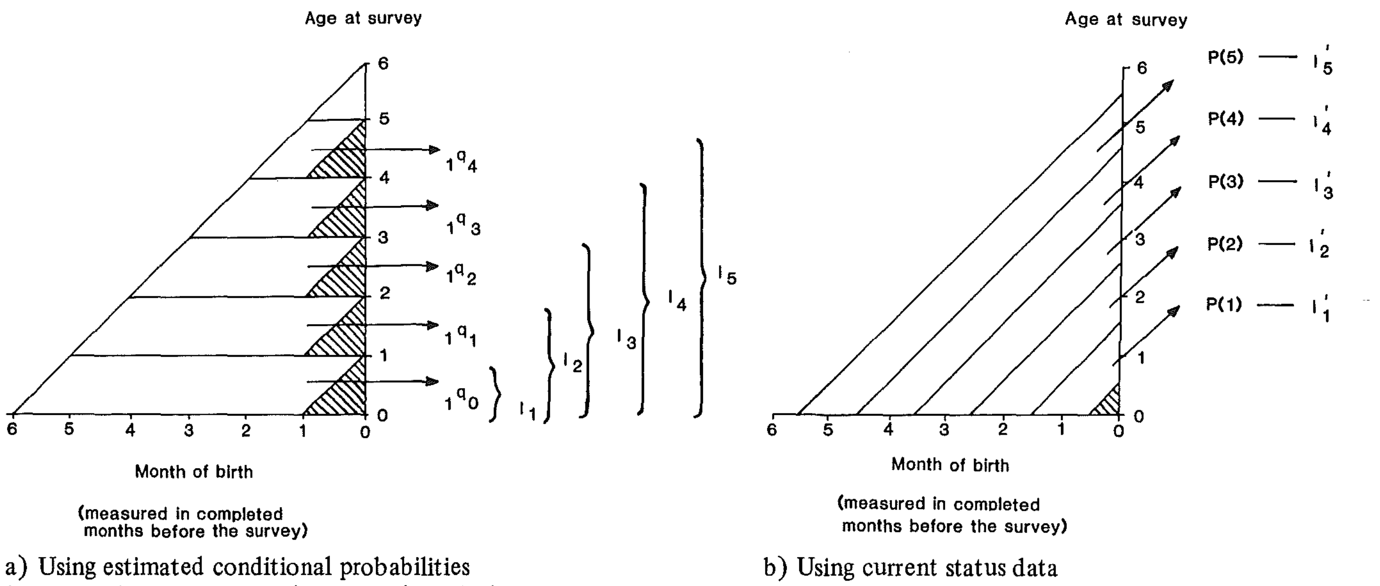


Figure 11 Reported, P(d), and standard, $P_s(d)$, proportions still being breastfed by the number of months elapsed since the birth (logit scale): surviving children (PFS)



a) Using estimated conditional probabilities incorporating retrospectively reported survival times

b) Using current status data

Figure 12 Comparison of the contribution of each reported birth cohort to the estimated survivor function using current status data and using estimated conditional probabilities

If we denote the total number of children currently breastfed (irrespective of their age) by B, and the average number of births per month by $\bar{N}$, then we can estimate the mean duration of breastfeeding, in months, as:

$$\bar{Y} = B/\bar{N} \qquad (14)$$

For any characteristic the prevalence (here the observed number currently breastfeeding) is a function of the incidence (the number who start breastfeeding) and the duration of the characteristic (here still breastfeeding). With a constant stream of entrants, the mean duration can be estimated by dividing the observed prevalence by the estimated incidence. Note that by defining the denominator in equation (14) as the monthly number of births rather than as the monthly number of children who start breastfeeding (the incidence of breastfeeding), the necessary allowance for those who are never breastfed has already been built in to our estimate.[36]

Equation (14) provides a remarkably simple means of estimating the mean duration of breastfeeding. It is also rather robust, because it is relatively insensitive to errors in the reported dates of birth for the children in question. The numerator does not require any information on dates at all. The denominator does, but it can be estimated from births in the year preceding the survey, or from the 2 years, or the 3 years, etc. Any misreporting of dates will affect the result only if the misreporting transfers births across the boundary of the period chosen. We are free to choose the period that we think minimizes any such transfers (we are not obligated to use the Z years preceding the survey). A period of one year is often too short (especially if we want to make estimates for subgroups) because of relatively small sample sizes. In Pakistan in addition, a period of one year did not appear advisable because of the apparent telescoping of dates and overstatement of births in the year preceding the survey. We have, therefore, used a period of two years. To be more specific, taking into account the fact that dates are available as calendar month differences (which give for a group of children apparently born 24 months before the survey some children born 23 completed months before and some born 24 completed months) and taking also into account the fact that there is noticeable heaping on the figure 24 months, we took as the total number of births that occurred in the two years before interview all those for whom the difference between calendar month of interview and calendar month of birth was 0–23, plus one-half those for whom this difference was 24 months.

Figure 13 compares the means estimated this way with those estimated by the preceding method for several major subgroups. The results are, in general, very consistent, suggesting that if we are only interested in the mean, the short method does as well as the longer one. It seems to do even slightly better in Pakistan. We can see that the estimates obtained by the short method yield results that are rather systematically 1–2 months higher than those yielded by the preceding method, suggesting that our choice of a two-year period for estimating the number of births per month has reduced the probable downward bias in the preceding estimates resulting from the presumed overstatement of births in the last year. Also, we find that within an age or education category the present estimates systematically rise with age and fall with education (there being only one exception – educated women relative to uneducated women in the youngest age group), whereas with the preceding estimates there were several exceptions.

Clearly, if we are interested only in the mean duration and *if it is reasonable to assume that the number of births per month has been roughly constant*, then this a very powerful, simple method. We have used it to examine the mean duration of breastfeeding for Pakistan as a whole and for its major socio-economic subgroups.

4.3 PRINCIPAL RESULTS

National Estimates

Table 3 summarizes the main results at national level for births in the four years preceding the survey in terms of the proportion ever breastfed and the mean duration of breastfeeding. More detailed tables of the estimated 'survivor' function can be found in appendix A.

In Pakistan, breastfeeding is both nearly universal and prolonged. 94.3 per cent of all children born in the four years before the survey were breastfed; the average duration of breastfeeding (including a count of zero months for those children who were never breastfed) was 19.2 months. For a large proportion of those who were never breastfed or were breastfed for only a very short period, breastfeeding was probably prevented or truncated by an early infant death.[37] If we look at children who survived to the survey, we find that over 98 per cent were breastfed and that the average duration of breastfeeding was 21.8 months.

Age of the child's mother at the time of the birth does not seem to have much impact on whether or not the child was breastfed, although those born to older women may have had a slightly higher chance of being breastfed. The average duration of breastfeeding, however, varies more markedly: children born to women 15–24 years of age were breastfed about 5 months less on average than children born to women aged 35–49. The latter were breastfed on average for a period very close to the two years traditionally favoured by Islam,[38] the former for several months less.

---

[36] We can also note that the sum of the proportions used on p26 $\Sigma P(d) = \Sigma [B(d)/N(d)]$ reduces to $\Sigma P(d) = \Sigma [B(d)]/\bar{N}$ if N(d) is constant. Note that by replacing $\Sigma B(d)$ in the numerator by B, the total number being breastfed regardless of their date of birth, we have effectively built in an allowance for those born in the calendar month of interview for whom we had to introduce an additional term earlier.

[37] It is very difficult to evaluate formally the effect of infant mortality on breastfeeding, partly because of the possibility of effects running in either direction and partly because the Pakistan standard recode data tape does not lend itself to it (we cannot easily use full life-table methods treating death as a competing risk because the age at death is coded only in rather broad categories and also because the reported age at death and the reported duration of breastfeeding are not infrequently inconsistent).

[38] The following quotation from the Koran refers explicitly, though non-directively, to a period of two years: 'And we have enjoined on man (to be good) to his parents: in travail upon travail did his mother bear him and in years twain was his weaning' (Sura 31:14). More directively one finds: 'The mothers shall give suck to their offspring for two whole years, if the father desires to complete the term' (Sura 2:233). (English-language version in Abdullah Yusuf Ali (1946), *The Holy Quran: Text, Translation and Commentary*, McGregor and Werner, pp 93 and 1083).

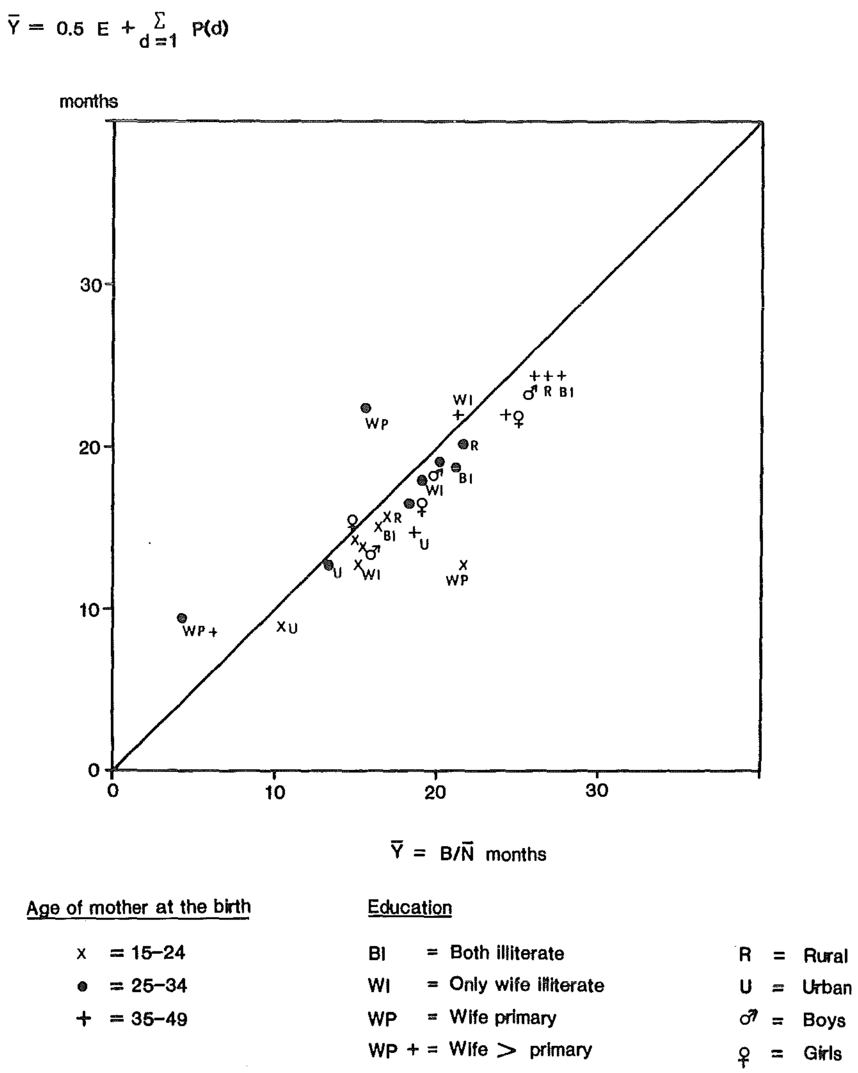


Figure 13 Comparison of two sets of estimates of the mean duration of breastfeeding derived from current status data: surviving children born in the four years preceding the survey (PFS)

Table 3 Estimates of breastfeeding for children born in the four years preceding the survey, by survival status of child and by age of mother at the birth (PFS)

	% breastfed	Mean duration of breastfeeding (months)	Total births in the 4 years
All children			
All women	94.3	19.2	4461
Women aged: 15–24	92.4	17.6	1917
25–34	95.5	19.7	1920
35–49	96.5	22.7	590
Surviving children			
All women	98.5	21.8	3761
Women aged: 15–24	97.7	20.4	1570
25–34	98.9	22.0	1658
35–49	99.4	25.9	507

NOTE: Mean duration of breastfeeding in months was estimated as $B/\bar{N}$, where B = number of children currently being breastfed at time of the survey, and $\bar{N}$ = average number of children born per month.

Table 4 Estimates of breastfeeding for children born in the four years preceding the survey, by sex of child, by survival status of child and by age of mother at the birth (PFS)

	Sex of child	% breastfed	Mean duration of breastfeeding (months)	Total births in the 4 years
All children				
All women	Boys	93.9	19.5	2270
	Girls	94.8	18.9	2189
Women aged 15–24	Boys	91.9	17.8	999
	Girls	93.0	17.4	924
Women aged 25–34	Boys	95.5	20.2	930
	Girls	95.5	19.4	989
Women aged 35–49	Boys	95.6	23.1	328
	Girls	97.7	22.3	262
Surviving children				
All women	Boys	98.5	22.3	1934
	Girls	98.5	21.4	1828
Women aged 15–24	Boys	97.8	20.7	816
	Girls	97.7	20.0	754
Women aged 25–34	Boys	98.8	22.5	816
	Girls	99.1	21.5	841
Women aged 35–49	Boys	99.5	26.4	286
	Girls	99.2	25.1	222

NOTE: Mean duration of breastfeeding in months was estimated as $B/\bar{N}$ where B = number of children currently being breastfed at time of the survey and $\bar{N}$ = average number of children born per month.

Table 5 Estimates of breastfeeding for children born in the four years preceding the survey, by rural-urban residence of mother, by survival status of child and by age of mother at the birth (PFS)

	Residence	% breastfed	Mean duration of breastfeeding (months)	Total births in the 4 years
All children				
All women	Rural	95.1	20.3	3258
	Urban	92.3	16.3	1202
Women aged 15–24	Rural	93.6	18.5	1400
	Urban	89.4	15.2	516
Women aged 25–34	Rural	96.1	20.9	1376
	Urban	94.0	16.7	544
Women aged 35–49	Rural	96.4	23.5	453
	Urban	97.1	19.4	139
Surviving children				
All women	Rural	99.4	23.0	2732
	Urban	96.3	18.7	1030
Women aged 15–24	Rural	99.0	21.4	1132
	Urban	94.6	17.7	438
Women aged 25–34	Rural	99.6	23.3	1188
	Urban	97.3	18.6	469
Women aged 35–49	Rural	99.7	26.6	390
	Urban	98.3	23.5	117

NOTE: Mean duration of breastfeeding in months was estimated as $B/\bar{N}$, where B = number of children currently being breastfed at time of the survey, and $\bar{N}$ = average number of births per month.

As table 4 shows, the Pakistan Fertility Survey data provide no evidence of a marked differential in breastfeeding by sex of the child. Excluding children who died, there is no systematic difference in terms of the proportions breastfed. The estimated average duration of breastfeeding, however, is systematically about one month longer for boys than for girls.

Estimates for Some of the Major Subgroups

We can also calculate estimates of summary statistics for some of the major subgroups in the population. Table 5, for example, compares births to women currently residing in rural areas to births to women residing in urban areas;

Table 6 Estimates of breastfeeding for children born in the four years preceding the survey, by education of mother, by survival status of child and by age of mother at the birth (PFS)

	Education	% breastfed	Mean duration of breastfeeding (months)	Total births in the 4 years
All children				
All women	Both illiterate	95.2	20.1	2493
	Only wife illiterate	94.5	18.8	1484
	Wife primary	91.5	18.2	316
	Wife > primary	84.5	11.2	168
Women aged 15–24	Both illiterate	93.5	18.2	986
	Only wife illiterate	92.8	17.6	686
	Wife primary	91.0	20.9	156
	Wife > primary	(81.3)	(10.5)	90
Women aged 25–34	Both illiterate	96.1	20.6	1117
	Only wife illiterate	95.9	19.6	614
	Wife primary	93.0	17.9	130
	Wife > primary	(88.8)	(12.2)	77
Women aged 35–49	Both illiterate	97.0	23.7	392
	Only wife illiterate	97.7	20.7	169
	Wife primary	–	–	26
	Wife > primary	–	–	2
Surviving children				
All women	Both illiterate	99.2	23.2	2080
	Only wife illiterate	98.8	21.3	1258
	Wife primary	97.2	19.8	274
	Wife > primary	89.7	12.3	150
Women aged 15–24	Both illiterate	98.6	21.3	788
	Only wife illiterate	98.0	20.3	570
	Wife primary	97.7	19.8	134
	Wife > primary	(87.2)	(11.5)	78
Women aged 25–34	Both illiterate	99.4	23.3	942
	Only wife illiterate	99.1	21.5	532
	Wife primary	98.2	18.6	113
	Wife > primary	(92.3)	(15.4)	71
Women aged 35–49	Both illiterate	99.8	27.1	333
	Only wife illiterate	100.0	23.4	149
	Wife primary	–	–	24
	Wife > primary	–	–	1

NOTES:
1 Mean duration of breastfeeding was estimated as $B/\bar{N}$, where B = number of children currently being breastfed at time of survey, and $\bar{N}$ = average number of births per month.
2 Small sample estimates are indicated as follows:
( ) = groups with less than 100 births in last 4 years. – = groups with less than 50 births in last 4 years.

table 6 compares the estimates for children born to women with different levels of exposure to education.

The comparison of rural and urban women shows that children born to rural women have a significantly higher chance of being breastfed. Regardless of whether or not we include children who died before the survey, 4–5 per cent fewer were breastfed among children born to urban women aged 15–24, and 2 per cent fewer among those born to urban women aged 25–34 (for births to older mothers the urban sample is rather small but suggests little difference between and rural and urban mothers).

In all age groups, the average duration of breastfeeding is systematically shorter – by 3.5 months – among children born to urban women.

The comparison of estimates broken down by educational exposure of the mother suggests only a small decline in breastfeeding (about two months) as one moves from the lowest education level considered here (neither husband nor wife literate) to women with primary education. The estimates for children born to women with education beyond the primary level are based on rather small samples, but they do suggest *markedly* less breastfeeding in this group.

We might continue along these lines making more detailed breakdowns, but we would increasingly run into serious problems of small sample sizes as the sample was split up further and further. Moreover, it is of some importance to examine the impact of more than one variable at a time. We shall, therefore, turn now to the whole question of multivariate analysis of the data.

5 Analysis of Differentials: Multivariate Analysis

The number and range of socio-economic, psychological and cultural variables included in most WFS questionnaires is too limited to test a comprehensive model of the determinants of breastfeeding behaviour, but one can still make good use of multivariate analysis techniques to examine and describe the differentials associated with those variables that are available. In this paper we shall simply analyse the differentials associated with a few basic socio-economic variables that are very widely available.[39] The methods can easily be extended to other variables.

Two major questions have to be dealt with. First, what data set should be used? Secondly, which analysis techniques would be most appropriate for treatment of the chosen data set?

5.1 CHOICE OF A DATA SET

More often than not the order of these two questions has been reversed, and researchers have tended to choose the retrospectively reported data from the last closed birth interval because these data can be analysed using familiar techniques. Breastfeeding has then usually ended for every interval considered, so there are no censored cases; we have a continuous variable to use as our dependent variable, that is conveniently measured in months. Most classic forms of regression analysis would then apply and the potential problems associated with censored data would be avoided. The unfortunate aspect of this approach, however, is that retrospectively reported data for the last closed birth interval are not only heaped,[40] but perhaps more importantly constitute a biased data set. In other words, ease of analysis is achieved at the expense of analysing a data set with a significant potential for being both distorted by selection biases and poorly reported.

It is tempting in multivariate analysis to ignore selection biases by invoking the hypothesis that the data for all subgroups are biased to essentially the same extent and that the differences between subgroups will, therefore, not be significantly affected. But this is often more a hope behind which one is tempted to seek shelter from the need to use other data sets and less familiar analytical techniques than a tenable hypothesis. For example, we know that selection biases in data for the last closed birth interval are most severe among subgroups with short marriage durations or long interbirth intervals, and we cannot expect the bias to be the same for all socio-economic subgroups. That the biases may sometimes differ to a significant extent has been suggested by a simple experiment (Page *et al* 1980). Analyses of differentials by socio-economic group in the Pakistan Fertility Survey data were shown to yield different results depending upon which data set (birth set) was used.[41] An attribute associated with noticeably longer than average breastfeeding in one data set did not necessarily figure so prominently or was even associated with relatively short breastfeeding according to another set.[42]

Clearly we should analyse a set of births that is not subject to marked selection biases. As before, our personal preference goes to the analysis of all births in a given period immediately preceding the survey, and to analyses that give each birth equal weight. Note that the results then reflect differences in how children born into different circumstances (eg born to mothers with different characteristics) fare. Again we should point out that by assigning different weights to the births on the basis of the number of children the mother concerned had given birth to in the period, the same data set could be used for an analysis in which each woman was given equal weight. The result would then estimate differences between the behaviour of *women* with different characteristics rather than between the experience of children, but would of course refer only to women who had given birth at least once during the period concerned.

5.2 CHOICE OF ANALYTICAL STRATEGY

Multivariate analysis deals with (i) the problem of discovering a structure in the data set related to a set of selected covariates, and (ii) the problem of efficiently estimating the parameters that describe this structure. Efficiency in estimating these parameters can be enhanced to the extent that the analyst already has a clear idea of the form of the underlying structure, as it is then possible to justify the use of a clearly defined statistical model. The more covariates we want to examine or the more categories we want to distinguish for each covariate, the more important the model becomes: the number of parameters soon becomes too large to be estimated from the available data unless we

[39] See the analysis for Thailand (Knodel and Debavalya 1980) and for Sri Lanka (Akin *et al* 1981) for uses of other variables. Popkin *et al* (1979) present a detailed conceptual framework for studying breastfeeding behaviour.

[40] Digit preference in retrospective reports (or in the information on age of the child concerned, which is used in analysis of current status data) can be partially overcome by the use of robust parametrization techniques (such as identification of the best-fitting model schedule), provided sample sizes are not too small. Akin *et al* (1981) suggest another method for reducing the effects of digit preference in probit analysis.

[41] Breastfeeding differentials were estimated using multiple classification analysis for each of three data sets: retrospective data for the last closed birth interval for each woman, current status data for each woman's current open birth interval, and current status data for all births in the four years preceding the survey. See section 5.2 for a discussion of the limitations of multiple classification analysis in this context.

[42] Of the three data sets considered, the analysis based on the last closed birth interval was the odd man out: the other two data sets yielded results that were closer to each other than to the results from the last closed birth interval.

can assume that a particular statistical model is an adequate approximation to at least part of the underlying structure.

The discovery of simple structures among a limited number of variables can often be achieved with rather simple tools if we use aggregate level data. A classic strategy consists of three stages. First the data are split into subgroups, each subgroup corresponding to one of the possible combinations of characteristics of the covariates. For each subgroup, one or more parameters describing the function of interest (here duration of breastfeeding) are then estimated (using procedures like those already used at national level in chapter 4, which allow for censored cases). Finally the resulting set of parameter estimates is examined and their basic structure extracted and summarized. The strategy is simple, but it is not very efficient because it is based on estimates for aggregates. As the number of covariates or the number of their categories increases, the number of possible combinations of categories and the corresponding number of subgroups needed rises rapidly, resulting in sample fragmentation, and the subgroups soon become too small to permit reliable estimation of their parameters. However, where sample fragmentation is not too severe the strategy not only gives the analyst a good first grasp of the data and a preliminary idea of what assumptions can be made (and which not) in any subsequent analysis, but it also gives first estimates of some of the effects. This type of exploratory analysis is presented in section 5.3.

With the insights gained, we are often in a better position to choose the underlying model that must be assumed if we are to use techniques like multiple regression to estimate the effects more efficiently. At this point we may also want to use individual rather than aggregate level data in order to carry our analysis as far as possible.[43] But censoring can cause serious problems for the classic forms of regression analysis. The only information available for all the cases without exception is current status (a 0–1 variable), which we might try to analyse with duration of exposure included as an additional regressor; expressed another way, we might look at the proportions still breastfeeding, controlling for duration of exposure. Use of a 0–1 individual level variable, or a proportion, in ordinary least squares regression can lead to two problems. First, it is possible for the proportions predicted by the fitted equation to take impossible values – outside the range (0, 1) – for some combinations of the covariates. Secondly, the error structure may be quite markedly heteroscedastic, complicating attempts to make significance tests for the estimated effects.

Classic ordinary least squares regression procedures (including dummy variable regression and multiple classification analysis) usually give adequate results when the proportions fall within the range 0.2–0.8.[44] Where proportions outside this range occur, the results become increasingly suspect. Life-table types of phenomena (ie all phenomena where the time that elapses before a particular event occurs is variable and of essential importance) usually include proportions outside the 'safe' range.

One standard way of handling proportions is to constrain values within the range (0, 1) by working with logit or probit transforms of the proportions rather than with raw proportions. Another set of methods is oriented explicitly towards life-table types of phenomena, rather than to proportions in general, and tends to be based not on proportions observed at each duration of exposure (the survivor function) but on the hazard rate estimated for each duration;[45] proportional hazards life-table models and related methods fall in this group. Both types of procedure tend, however, to require more software and also more computer space than are widely available, especially in developing countries. Moreover, experience with their application to breastfeeding patterns is still limited.

It is possible though to use widely available and familiar classic regression programs with slight modifications that reduce, even if they do not eliminate, the problems. The procedures are quite easy to apply but slightly *ad hoc.* They reduce the problems considerably, however, although they do not tackle them at their root.

Given the dilemma that the preferred techniques are not everywhere available, we shall give here first a discussion of one simple way of applying widely available and familiar classic regression programs; we shall then present more briefly the results from two of the more thorough-going procedures, use of logit-linear regression and of proportional hazards models.[46] The choice between the various methods for any particular analysis will depend heavily on the facilities available. For Pakistan, at least, they yield broadly similar results, although some differences do emerge.

5.3 PARAMETRIZATION OF BREASTFEEDING TABLES FOR SUBGROUPS

Since the approach based on analysis of estimated parameters of the breastfeeding tables for each subgroup separately is the most exploratory and also the most transparent, we shall examine this first.

We could carry out this type of analysis using any of the various characteristics we have estimated in chapter 4. If we are most concerned with a single summary statistic describing just one aspect of breastfeeding patterns, we might well simply choose a measure of central tendency, such as the mean or median duration of breastfeeding; if we are interested in characterizing the entire function, we might use α and β. Since first interest often goes to measures of central tendency and since estimation of the parameters needed to reproduce the entire function is less reliable (β in particular is hard to estimate reliably from small samples or data with strong heaping), we shall focus here on analysis of the mean

---

[43] The aggregate level structure and individual level structure are, of course, not necessarily the same.

[44] Some authors set more restricted limits, recommending rejection of ordinary least squares regression when there are proportions outside the range 0.25–0.75 (see, for example, Namboodiri *et al* 1975:345).

[45] The hazard rate at duration d is the probability of experiencing the event in question (here weaning) in the next instant, conditional or not on having experienced it before d. The 'force of mortality' is the hazard rate for mortality, for example. It can be approximated as (ln $_nw_d/n$), where $_nw_d$ is the conditional probability of experiencing the event between durations d and (d + n), ie $_nw_d = (P(d) - P(d + n))/P(d)$ where P(x) is the proportion who have not yet experienced the event at duration x.

[46] The reader interested in actually applying proportional hazards models should consult McDonald (1981) and the references cited therein (especially Cox (1972), Kay (1977) and Menken *et al* (1981)). For an example of the application of probits to breastfeeding data, see Akin *et al* (1981).

Table 7a Median polish on the estimated mean durations of breastfeeding for children of rural women (PFS)

(a) Original table

Ed. \ Age	15–24	25–34	35–49
BI	21.9	23.9	27.4
WI	21.0	22.4	23.5
WP	21.0	19.6	–
W > P	–	–	–

(b) Extraction of preliminary overall median

15–24	25–34	35–49	
–0.3	+1.7	+5.2	
–1.2	+0.2	+1.3	
–1.2	–2.6	–	
–	–	–	
			22.2

1st round

(c) Extraction of col. medians from (b)

+0.9	+1.5	+1.9	
0.0	0.0	–2.0	
0.0	–2.8	–	
–	–	–	
–1.2	+0.2	+3.3	22.2

(d) Extraction of row medians from (c)

–0.6	0.0	+0.4	+1.5
0.0	0.0	–2.0	0.0
+1.4	–1.4	–	–1.4
–	–	–	–
–1.4[a]	0.0	+3.1	22.4[b]

2nd round

(e) Extraction of col. medians from (d)

–0.6	0.0	+1.2	+1.5
0.0	0.0	–1.2	0.0
+1.4	–1.4	–	–1.4
–	–	–	–
–1.4	0.0	+2.3	22.4

(f) Extraction of row medians from (e) (estimates are here seen to have stabilized)

–0.6	0.0	+1.2	+1.5
0.0	0.0	–1.2	0.0
+1.4	–1.4	–	–1.4
–	–	–	–
–1.4	0.0	+2.3	22.4

Final table

Ed. \ Age	15–24	25–34	35–49	
	Residuals			Educ. effects
BI	–0.6	0.0	+1.2	+1.5
WI	0.0	0.0	–1.2	0.0
WP	+1.4	–1.4	–	–1.4
W > P	–	–	–	–
	Age effects			Median[c]
	–1.4	0.0	+2.3	22.4

[a]Estimate obtained in (c) row median in (c), ie (–1.2 – (+0.2)).
[b]Estimate obtained in (c) + row median in (c), ie (+22.2 + (+0.2)).
[c]Median standardized on a uniform distribution by age and education.
NOTE: – = sample size too small for satisfactory estimation of mean duration.

durations for the various subgroups.[47] Essentially the same procedures could be used, however, for any of the parameters.

We have already shown in chapter 4 that the mean duration of breastfeeding in Pakistan differs somewhat according to age, education and rural-urban residence of the mother. The estimates can conveniently be displayed in a three-way table (table 7). In order to examine the structure and to quantify the effect of each variable on the mean duration of breastfeeding, we have used the 'median polish' procedure proposed by Tukey (1977). In essence this is based simply on the extraction of a value representing the overall level of the estimates in the table and the estimation of the extent to which each of the categories of the covariates (represented by the various rows and columns in the table) deviate from this, using in the first place a simple additive model. For each two-way table we have:

[47]We shall use here the estimated mean duration defined by $B/\bar{N}$ where B is the number of children currently being breastfed and $\bar{N}$ the estimated number of births per month.

Table 7b Median polish on the estimated mean durations of breastfeeding for children of urban women (PFS)

(a) Original table

Ed. \ Age	15–24	25–34	35–49
BI	19.0	20.7	23.4
WI	18.8	19.2	22.6
WP	18.9	17.9	–
W > P	12.2	13.4	–

(b) Extraction of preliminary overall median[a,b]

15–24	25–34	35–49	
–0.1	+1.6	+4.3	
–0.3	+0.1	+3.5	
–0.2	–1.2	–	
–6.9	–5.2	–	
			19.1[b]

1st round

(c) Extraction of col. medians from (b)[b]

+0.1	+1.5	+0.4	
–0.1	0.0	–0.4	
0.0	–1.3	–	
–6.7	–5.3	–	
–0.2[b]	+0.1[b]	+3.9[b]	19.1[b]

(d) Extraction of row medians from (c)

–0.3	+1.1	0.0	+0.4
0.0	+0.1	–0.3	–0.1
+0.6	–0.7	–	–0.6
–0.7	+0.7	–	–6.0
–0.3[b]	0.0[b]	+3.8[b]	19.2[b]

2nd round

(e) Extraction of col. medians from (d)

–0.3	+1.0	+0.2	+0.5
0.0	0.0	–0.1	0.0
+0.6	–0.8	–	–0.5
–0.7	+0.6	–	–6.0
–0.3[b]	+0.1[b]	+3.6[b]	19.1[b]

(f) Extraction of row medians from (e)

–0.5	+0.8	0.0	+0.7
0.0	0.0	–0.1	0.0
+0.7	–0.7	–	–0.6
–0.6	+0.7	–	–6.0
–0.4[b]	0.0[b]	+3.5[b]	19.2[b]

3rd round

(g) Extraction of col. medians from (f) (estimates are here seen to have stabilized)

–0.5	+0.8	0.0	+0.7
0.0	0.0	–0.1	0.0
+0.7	–0.7	–	–0.6
–0.6	+0.7	–	–6.0
–0.4[b]	0.0[b]	+3.5[b]	19.2[b]

Final table

Ed. \ Age	15–24	25–34	35–49	
	Residuals			Educ. Effects
BI	–0.5	+0.8	0.0	+0.7
WI	0.0	0.0	–0.1	0.0
WP	+0.7	–0.7	–	–0.6
W > P	–0.6	+0.7	–	–6.0
	Age effects[b]			Median[a,b]
	–0.4	0.0	+3.5	19.2

[a]Median standardized on a uniform distribution by age and education.
[b]Median and col. effects calculated on basis of first 3 rows only (ie excluding W > P) since 4th row is not represented in the corresponding rural tables.
NOTE: – = sample size too small for satisfactory estimation of mean duration.

Observed value = Overall value + Row effect + Column effect + Residual

or $O_{ij} = A + R_i + C_j + \epsilon_{ij}$

where O_{ij} is the observed value for the subgroup in the ith row and the jth column.

A is the overall level of values in the table

R_i is the effect of the category represented in the ith row

C_j is the effect of the category represented in the jth row, and

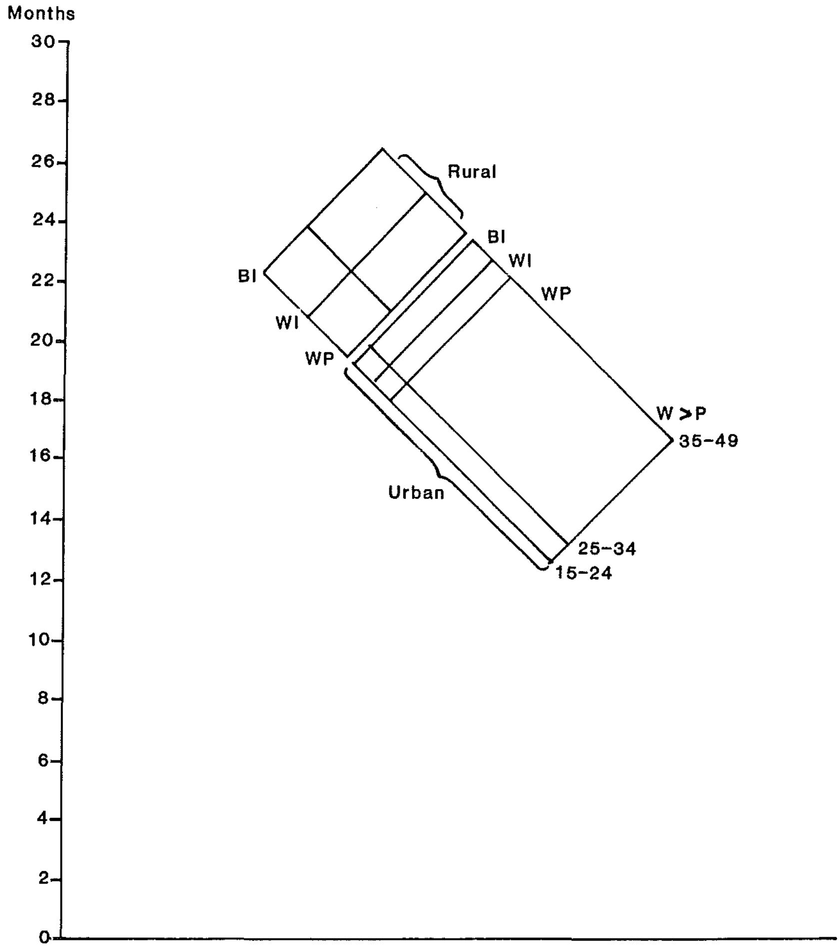
ϵ_{ij} is the residual value in cell ij.

Since the analysis is exploratory (an additive model may not be appropriate), we chose to use medians rather than the customary arithmetic means to estimate A, R_i and C_j. Use of medians has the advantage of drawing attention to cells that have atypical values in the context of the initial model (the absolute values of the residuals, or their squares, are not minimized as they would be if we used arithmetic means. Inappropriateness of the model will thus show up more easily in the form of large residuals or a distinct pattern in the residuals. Use of medians has the additional advantage that the estimates of effects are less easily distorted by the occasional wild values that can result from measurement problems (eg cell values based on small samples).

Tables 7a and 7b illustrate how the effects are estimated for each two-way table, once a preliminary median has been extracted from the original table and preliminary residuals calculated (sub-table (b)). The median of each column is extracted and the first-stage residuals calculated (c); the median of each row of these is then extracted and new residuals calculated (d). The process is repeated until the table stabilizes, ie until further repetitions do not change the values.

Missing values in just a part of any row or column (typically cells with small samples for which no estimates can be made) are handled automatically by this procedure. Where an entire row or column is missing for one sub-table and not for the other(s), however, we have to take special steps. In Pakistan, for instance, there are very few women with post-primary education in any age group in the rural areas. The row representing the highest education category is, therefore, entirely devoid of estimates in the rural sub-table. If we are to compare the results for the two sub-tables, they must both be analysed in the same way. Since the overall median and the column effects for the rural table must be estimated from just the first three rows, so they must be estimated from just the same three rows for the urban areas. In other words, we must exclude the highest education row when estimating the overall level and the column effects in the urban sub-table. Note that row effects can still be estimated for all rows that are not entirely blank (including, therefore, the fourth row in the urban sub-table), but all row effects are measured as deviations from the median of the first three rows, not from the median of all four rows. The final estimates are essentially standardized for the age and educational structure of the population with each combination of age and educational level (excluding post-primary education) being given equal weight.

The results are given at the bottom of tables 7a and b (estimated effects and residuals) and in figure 14 (fitted values for each subgroup). The overall age and education-standardized median duration of breastfeeding (excluding women with post-primary education) is immediately seen to be about three months longer (22.4–19.2) in rural than in urban areas. In both urban and rural areas, being born to a woman between 35 and 49 years of age adds about four months to the duration of breastfeeding, compared with being born to a woman in the age range 15–24. Deviations from the overall level related to education appear to be slightly smaller than those related to our three age



NOTE: Estimated values are those derived from the median polish summarized in tables 7a and 7b.

Figure 14 Estimated median duration of breastfeeding by age, education and urban-rural residence of the mother: surviving children born in the four years preceding the survey (PFS)

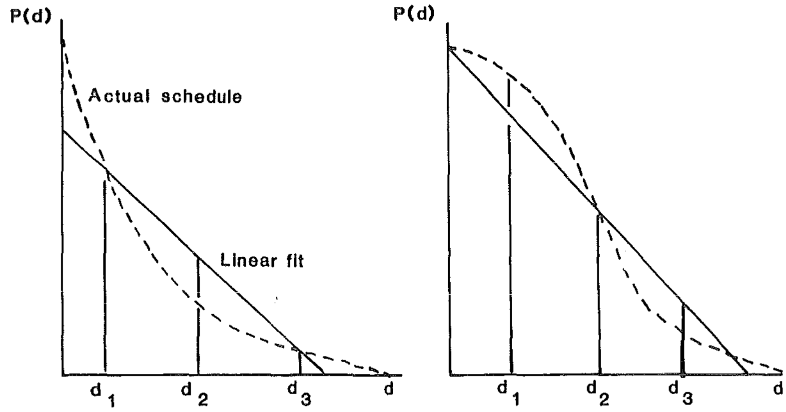


Figure 15 Types of errors in estimated proportions still breastfed introduced by the assumption of linearity

groups if we consider the first three education groups: the difference between the apparent effect of the couple being illiterate and that of the wife having had primary education is 1.5–3 months. Post-primary education, however, seems to be associated with a marked deviation relative to the other three educational categories: breastfeeding is reduced by some six months. The residuals are not negligible but neither are they extremely large; no residual is larger than the estimated difference between the effects of the two extreme categories of either age or education, and most are smaller than the estimated difference between adjacent categories. In addition, we can note that there is no particularly striking pattern in the residuals. Overall, we can conclude from the residuals that although the structure within each sub-table is not purely additive, an additive model probably yields reasonable first estimates of the impact of the various age and education categories in Pakistan. If we compare the two sub-tables, we may discern the suggestion of a slight interaction between rural-urban residence and the effects of age and education. The estimated effects of education appear to be slightly larger in rural than in urban areas; and the central age group seems to be closer to the youngest age group in its effect in urban than in rural areas. The differences are, however, small and could well fall within the margin of error of the data and the method.[48]

5.4 REGRESSION-TYPE ANALYSES USING INDIVIDUAL-LEVEL DATA

Application of Simple Regression Analysis Using Polynomials

We have already indicated that the estimated effects derived from the application of classic regression procedures to 0–1 variables are suspect because the predicted proportions are not constrained to respect the bounds 0 and 1. Moreover, even within these bounds, P(d) functions typically exhibit an S-shape that linear regression does not handle. The errors in the estimates that would be introduced by an assumption of linearity are of course a function of the schedule's shape. If P(d) has the shape shown in the left-hand panel of figure 15, the linear approximation is likely to give an adequate estimate of the proportion still breastfeeding for durations d1 and d3, but one would obtain a systematic overestimate at d2 and an intercept which would grossly underestimate the proportion initiating breastfeeding. If the schedule has the shape shown in the right-hand panel, systematic distortions would occur at durations d1 and d2.

The general shape can be captured quite easily by using a third degree polynomial (ie using d, d^2 and d^3 instead of just d) in the regression equation.[49] We can note that it is difficult to interpret the coefficients of d, d^2 and d^3 directly because of the perfect collinearity between them, but that if their interpretation is required the multicollinearity problem can be avoided by using orthogonal polynomials.

This procedure usually produces estimated proportions that decline monotonically and that are within the required bounds. However, it is still possible for small peculiarities to arise in the results. Peculiarities are especially likely to occur if important interaction terms are omitted, because the estimated regression coefficient of a particular regressor is then a weighted average of the real effects, which in fact vary across levels of the other regressors. Variables such as the birth cohort of the mother, educational level and area of residence are very likely to produce such interaction effects on breastfeeding as the following examples will bring out. A particular educational level, say primary schooling, is likely to have a different impact on the duration of breastfeeding depending on the birth cohort of the woman concerned or the region in which she lives. A woman of 40 living in a rural area who has primary education is much more exceptional in her environment than another woman with primary education who is only 20 and who lives in an urban area. In other words, a given educational level may mean very different things depending on the time and the environment in which it was received. Similarly, if different trend-effects with respect to breastfeeding are suspected in the various regions and if the birth cohort of the mother picks up a good portion of such a trend, an interaction effect between age of the woman and the area should emerge.

[48] We are after all, estimating effects by taking the median of just three rows (or three columns) at a time.

[49] David Smith (1981) has suggested an alternative way of modifying standard regression procedures to reduce the problems, which relates the current status of each child to the average value for all children born at the same time. The regression is carried out on the difference between individual and average current status.

Table 8 Specification of a regression model linking the proportion of children still being breastfed among those born in the 40 months preceding the survey to age, rural-urban residence and educational level of their mothers

Regression equation

$$PD = A + B_1 d + B_2 d^2 + B_3 d^3 + B_4 AM + B_5(d * AM) + B_6(d^2 * AM) + B_7(d^3 * AM) + B_8 BI + B_9(d * BI) + B_{10}(d^2 * BI)$$
$$+ B_{11}(d^3 * BI) + B_{12} WI + B_{13}(d * WI) + B_{14}(d^2 * WI) + B_{15}(d^3 * WI) + B_{16} WP + B_{17}(d * WP) + B_{18}(d^2 * WP)$$
$$+ B_{19}(d^3 * WP) + B_{20} U + B_{21}(d * U) + B_{22}(d^2 * U) + B_{23}(d^3 * U) + B_{24}(BI * AM) + B_{25}(WI * AM) + B_{26}(WP * AM)$$
$$+ B_{27}(U * AM) + B_{28}(BI * U) + B_{29}(WI * U) + B_{30}(WP * U)$$

Variable description		*Mean value*[a]	
d	= duration of exposure (ie age of child):	$\bar{X}$ = 19.5 months	
AM	= age of mother at time of child's birth:	$\bar{X}$ = 26.8 years	
BI	= dummy variable equal to 1 if both parents illiterate:	$\bar{X}$ = 55.1%	= 100%
WI	= dummy variable equal to 1 if only mother illiterate:	$\bar{X}$ = 33.9%	
WP	= dummy variable equal to 1 if mother has primary education:	$\bar{X}$ = 7.2%	
(W > P	= reference category, mother more than primary education):	($\bar{X}$ = 3.8%)	
U	= dummy variable equal to 1 if mother is living in urban area:	$\bar{X}$ = 26.9%	= 100%
(R	= reference category, mother living in rural area):	($\bar{X}$ = 73.1%)	
P(d)	= proportion of children still being breastfed at duration d:	$\bar{X}$ = 55.2%	
		N = 3224	

[a]The mean value for the dummy variable is simply the proportion with the given characteristic.
NOTES: the variables d, d^2 and d^3 can be replaced by three orthogonal polynomials to eliminate collinearity between them; children who died before the age of 2 years were excluded from this analysis.

A classic strategy to pick up interaction effects consists of running statistical tests whereby a full model (including the hypothesized interactions) is evaluated against the restricted model (without the interaction terms) (see Namboodiri *et al* (1975)). This type of strategy is most useful in the presence of numerous regressors. In such situations there may be so many interaction effects that estimation of all interaction terms would not be possible for lack of sufficient data points; a 'scouting expedition' is then necessary to reduce their number by picking up the most important ones. If, however, we have a more limited number of variables and sound substantive reasons for suspecting particular interactions, we can proceed directly to incorporate and estimate them. As we wanted to use the same variables in the present example as in the example with parametrization by subgroup in order to bring out similarities or differences in the methods, we shall simply proceed by including all first order multiplicative effects (ie within pairs of variables), while the higher order interactions are omitted.[50]

In fact, we shall find that some of the interaction terms are relatively unimportant and could have been omitted for a substantive analysis.

The regression model linking the proportion of children being breastfed (P(d)) to the length of exposure (d), the age of the mother at the birth of the child (AM), the educational categories (BI, WI, WP, W > P) and the residential categories (U, R) is given in table 8 together with the mean value for each of the variables. For this analysis we have included only births that occurred in the 40 months before the survey. This cut-off was used because nearly all subgroups reach zero as their proportion breastfeeding by this point and the estimates of the coefficients of the d terms beyond it become unsatisfactory. The values of the coefficients are given in table 9 where a cubic function in d is used, and in table 10 where orthogonal polynomials are used.[51] Both types of regression give the same fits but the coefficients in the set based on orthogonal polynomials are preferable for terms involving duration, as the problem of multicollinearity between d, d^2 and d^3 is eliminated. In both tables, the coefficients on the first line direct the general shape of the P(d) function while the other coefficients play a role in modifying this, depending on the value of the covariates. In table 9 for instance, the coefficient B_1 directs the general downward slope of P(d), B_2 produces an acceleration of the attrition due to weaning at intermediate durations, while B_3 produces a correction in the opposite direction at the tail of the P(d) distribution (ie at high values of d). The coefficients in the top line of table 10 can be interpreted in a similar way since the shape of each of the orthogonal polynomials is known. This shape depends only on the range of values used for d. In our example, with d ranging from 0 to 40, the first, OP1, is a linearly increasing function with a value of –20 at d = 0 and a value of +20 at d = 40; the second, OP2, is a parabola moving from +260 at d = 0 to a minimum of –140 at d = 20 and increasing again to +260 at d = 40; the third, OP3, moves from –2470 at d = 0 to +2181 at d = 11, falls back to 0 at d = 20 and further down to –1281 at d = 29 and moves up again to +2470 at d = 40. The negative coefficient B_1 turns the original linearly increasing effect of OP1 into a general downward slope for the P(d) function, the positive coefficient B_2 leaves the parabola effect of OP2 to emerge in the form of a steeper drop-off rate before d +

[50] Even if we consider only first-order interactions, combining the 3 terms in d with 1 variable for age of mother at the birth, with the 3 dummy variables needed to describe the impact of the 4 education groups and the 1 dummy variable needed to describe the impact of 2 residential categories, we come to no less than 22 interaction terms.

[51] The tables needed for the use of orthogonal polynomials are found in standard handbooks of statistical tables. See for example Beyer (1966): 366–79.

Table 9 Regression results of the model specified in table 8 with a third-degree polynomial in d (PFS)

		d	d^2	d^3
	73.7 (A)	−0.5922 (B_1)	−0.3381 (B_2)	+0.007633 (B_3)
AM	+0.675 (B_4)	−0.1063 (B_5)	+0.1268 (B_6)	−0.000265 (B_7)
BI	+12.1 (B_8)	+6.1436 (B_9)	−0.3032 (B_{10})	+0.003819 (B_{11})
WI	+14.2 (B_{12})	+6.1064 (B_{13})	−0.3204 (B_{14})	+0.004301 (B_{15})
WP	+14.0 (B_{16})	+5.7739 (B_{17})	−0.2971 (B_{18})	+0.003704 (B_{19})
U	+12.6 (B_{20})	−0.5044 (B_{21})	−0.0239 (B_{22})	+0.001003 (B_{23})

	BI	WI	WP	U
AM	−0.327 (B_{24})	−0.428 (B_{25})	−0.513 (B_{26})	−0.185 (B_{27})
U	−10.0 (B_{28})	−8.7 (B_{29})	−8.7 (B_{30})	

$R = 0.768$ $R^2 = 0.590$
N = 3224

NOTES d = duration of exposure (ie age of child)
AM = both parents illiterate
WI = only mother illiterate
WP = mother had primary education
U = urban

Table 10 Regression results of the model specified in table 8 with orthogonal polynomials (PFS)

		OP1	OP2	OP3
	4.48 (A)	−3.0329 (B_1)	+0.1199 (B_2)	+0.009160 (B_3)
AM	+1.052 (B_4)	+0.0163 (B_5)	−0.0032 (B_6)	−0.000318 (B_7)
BI	+33.83 (B_8)	−0.4422 (B_9)	−0.0741 (B_{10})	+0.004584 (B_{11})
WI	+33.83 (B_{12})	−0.4647 (B_{13})	−0.0623 (B_{14})	+0.005161 (B_{15})
WP	+35.19 (B_{16})	−0.3322 (B_{17})	−0.0648 (B_{18})	+0.004445 (B_{19})
U	+6.03 (B_{20})	−0.0027 (B_{21})	+0.0363 (B_{22})	+0.001204 (B_{23})

Regression coefficients B_{24} through B_{30} and values of R and R^2 are identical to those in table 9

N = 3224

NOTES: OP1, OP2, OP3 = orthogonal polynomials replacing d, duration of exposure (ie age of child), d^2 and d^3.
AM = age of mother at time of child's birth
BI = both parents illiterate
WI = only mother illiterate
WP = mother had primary education
U = urban

20 and a slower drop-off rate after d = 20, while the positive coefficient B_3 allows OP3 to give the P(d) function its final ogive-shape. Again, the coefficients of the interaction terms in which any of the three orthogonal polynomials appears together with another covariate modify this process. In this respect, the effect of the mother belonging to an older age cohort or having less than secondary education diminishes the parabolic effect of OP2 (see coefficients B_6, B_{10}, B_{14} and B_{18}), so that the dropping-off of the P(d) function is less steep before d = 20 than among children of younger women with more than primary education. The effect of urban residence, to the contrary, enhances the parabolic effect of OP2 (see coefficient B_{22}), which leads to a faster decline in the proportions still being breastfed before the duration of 20 months than among the children of rural women (reference category). Equally striking in table 17 is the fact that the three lower educational categories have nearly identical regression coefficients (compare the series B_8 to B_{11}, B_{12} to B_{15} and B_{16} to B_{19}). This implies that there is a marked distinction between them and the reference category, ie W > P, but that the internal differences between BI, WI and WP are small.

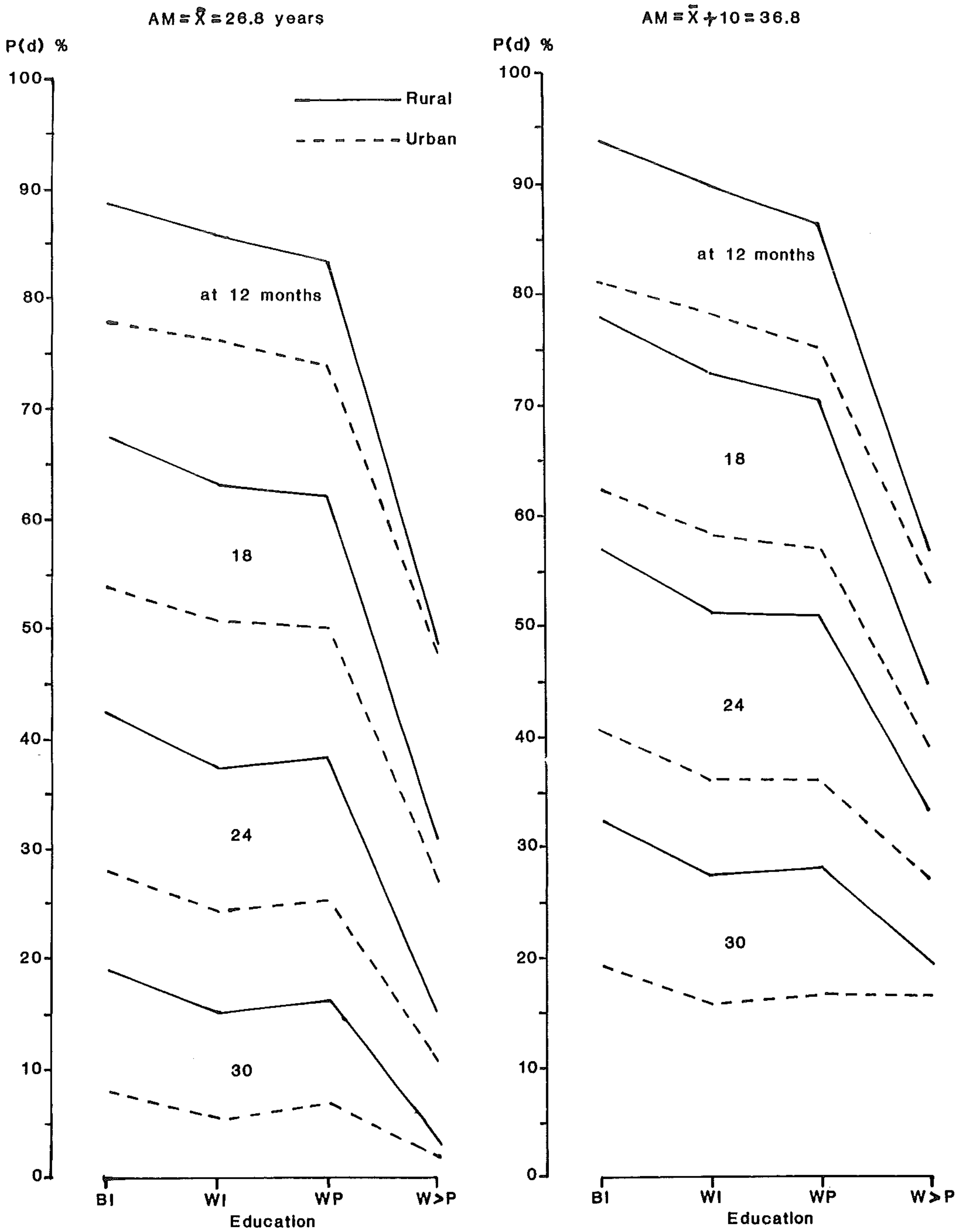
On the whole, it is not easy to interpret the coefficients of such intricate regression equations; one quickly needs a calculator to evaluate their numerical impact. The results can be used most profitably by producing descriptive tables or figures through which the effect of a particular covariate can be evaluated while fixing the values of the others. The results of such an application are shown in figure 16. The left-hand panel shows the proportions of children still being breastfed at ages 12, 18, 24 and 30 months by educational and residential categories of their mother, for children born to women aged 26.8. Since this is the mean age for the whole sample, these proportions are representative for the sample as a whole (or representative for the situation in which the age of mother was not introduced). The right-hand panel produces the results for children born to women ten years older.

The results are almost perfectly in line with those generated through the parametrization of the P(d)-function by subgroup: the education effect emerges in a very clear way only for women with more than primary education and the rural-urban differential increases as the mother's age increases.

Logit-Linear Regression

A standard procedure nowadays for analysing proportions is to apply a logit transform to them. This provides a convenient means for constraining predicted proportions. Because the transformed variable has a range $(-\infty, +\infty)$ corresponding to the range (0, 1) for a raw proportion, any predicted value whatever for the transformed variable implies a proportion within the range (0, 1). We can note in addition that the logit P(d) function is usually much closer to a straight line than is the original P(d) function.

The logit-linear model is an example of a generalized linear model with binomial error structure and logit link function. Suppose we organize our current status data on breastfeeding in a table format such that all categories of the independent variables of interest (including duration elapsed since the birth in question) are exhaustively classified against each other and that each cell shows the propor-



NOTE: Values of P(d) are estimated from the regression equation of table 8 and the coefficients reported in table 10.
AM = age of mother at birth of child
BI = both husband and wife illiterate
WI = wife only illiterate
WP = wife primary education
W > P = wife more than primary education (very few rural women have post-primary education).

Figure 16 Estimated proportions still being breastfed at ages 12, 18, 24 and 30 months, by age, education and urban-rural residence of the mother: regression results for surviving children born in the 40 months preceding the survey (PFS)

Table 11 Deviances from logit-linear regression models fitted to data for surviving children born in the 40 months preceding the survey (PFS)

Model[a]	Urban		Rural	
	df	Deviance[b]	df	Deviance
(1) DUR	564	498.1	543	487.6
(2) DUR, AGE	559	474.7[c]	538	445.6[c]
(3) DUR, AGE, ED	556	454.8[c]	535	436.6[d]
(4) DUR, AGE, ED, REG	554	454.5	533	423.1[d]
(5) DUR, AGE, (DUR X AGE)	472	373.6	449	340.5
(6) DUR, AGE, ED, REG (AGE X ED)	541	443.5	521	410.7
(7) DUR, AGE, ED, REG, (DUR X ED), (AGE X ED)	487	369.3	481	375.9
(8) DUR, AGE, ED, REG, (DUR X ED)	500	378.9[d]	492	381.8

[a]Interaction effects considered in a model are shown in parentheses.
[b]Chi-squared test was used to evaluate if the model fits significantly better than the preceding model (for models (2)–(4)). Model (5) was evaluated against model (2) and models (6)–(8) were evaluated against model (4) for the test of relative fit.
[c]Model fits significantly better at the 0.001 level.
[d]Model fits significantly better at the 0.05 level (0.01 level for model (4) in rural sample).

tion of children still being breastfed (P_c). Instead of dealing with the observed values of P (as in the case of the previous section using standard normal theory regression model), we use Y_c, where[52]

$$Y_c = \ln [P_c/(1 - P_c)] \quad (15)$$

In a model with logits as a dependent variable, the estimated effects imply the differences in logit value associated with the differences in the independent variables taken into account. We can of course compute the expected proportions implied by the fitted logit values by taking their antilogarithms. More specifically, if $\hat{Y}$ is the sum of the estimated effects from the model for cell C, the corresponding value of the proportion for cell c, ($\hat{P}_c$), can be obtained using equation (16) below:

Since $\hat{Y} = \ln [\hat{P}_c/1 - \hat{P}_c)]$,

we have $e^{\hat{Y}} = \hat{P}_c/(1 - \hat{P}_c)$,

and $$\hat{P}_c = e^{\hat{Y}}/(1 + e^{\hat{Y}}) \quad (16)$$

For our purpose, equation (16) should yield the expected proportions of children still being breastfed implied by the fitted model. We have run a logit analysis on the same data as in the preceding section using the GLIM package.[53] We should note that the two sets of results should not be expected to be exactly the same. There are slight differences in the fitting (the fit here being to the logit P(d) values rather than to the P(d) values themselves) and in the assumption of a binomial rather than a normal error structure. There are also differences arising from practical limitations. GLIM keeps all the information on the number of observations, number of independent variables and the parameters in the model in core, which required more space than was available to us. To reduce space requirements for the data we had to analyse births to rural and urban women separately. In addition, we had to use grouped categories for duration elapsed since the birth and for age of mother; more specifically, we collapsed the duration since the birth from 41 single-month categories (0–40 months) to 19 categories, and age of mother at birth was recoded to refer to 6 categories: 15–19; 20–24; 25–29; 30–34; 35–39; and 40–49.[54] We retained all four educational categories used earlier for the urban sample but excluded W > P (post-primary education) for the rural sample because of the very small number of cases with this level of education in the rural subsample. We were, however, able to analyse one other variable with a very limited number of categories – region of residence.[55]

The space requirements also made us consider only a few interaction effects among the independent variables. Nevertheless, we were able to evaluate eight models. Results are shown in table 11. Whereas the addition of a variable (or an interaction term) always leads to a reduction in the deviance, we can evaluate whether such a reduction is statistically significant by subtracting the degrees of freedom (dfs) and the deviances associated with the two models being compared against the theoretical distribution of the chi-square. Starting from the first model with DUR alone, each subsequent model was evaluated against the preceding model, except model 5 (which was compared with model 2) and models 6–8 which were compared individually with model 4. For the rural subsample, the model with the additive effects of all four independent variables (model 4) seems to be the most parsimonious, compared with model 3 (with DUR, AGE and ED) for the urban subsample. We have,

[52] The ratio $P_c/(1 - P_c)$ is often referred to as the odds in favour of a particular status, ie a child still being breastfed in our case. Therefore the Y_c, ie $\ln[P_c/(1 - P_c)]$ values refer to the log-odds or logit in favour of children still being breastfed (the term log-odds ratio and logit are used interchangeably). The notion of odds is frequently used in common parlance. For example, we say 'the odds are 4 to 1 in favour' of a particular event if the probability of occurrence of that event is 0.8.

[53] Useful references for using GLIM include Braun (1980), Little (1978) and McDonald (1981). The GLIM package is distributed by Numerical Algorithms Group, 13 Banbury Road, Oxford OX2 6NN, England.

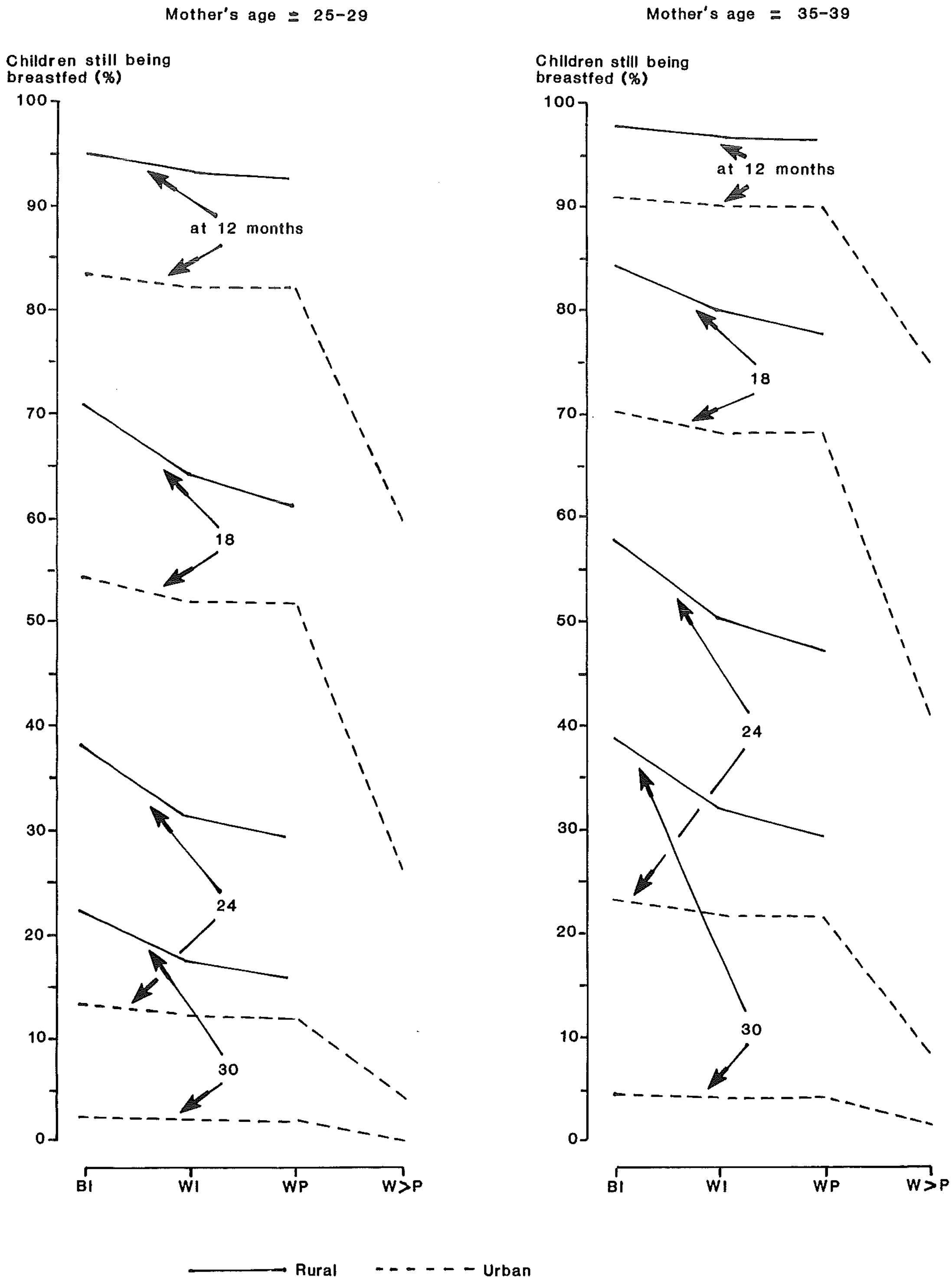
[54] Children born to women below age 15 were excluded from the analysis and so were children born to women of Baluchistan Province as we wanted to look into the regional differentials and the sample for Baluchistan was too small (below 10).

[55] It would of course have been possible to run separate analyses by region in the preceding analysis.

Table 12 Estimated proportions still being breastfed at ages 12, 18, 24 and 30 months, by age, place, province of residence and educational level of mother: logit-linear regression results for surviving children born in the 40 months preceding the survey (PFS)

Age of mother	Urban				Rural			
	Duration since birth (months)				Duration since birth (months)			
	12	18	24	30	12	18	24	30
A Both illiterate (BI)								
15–19	0.854	0.529	0.149	0.026	0.928	0.622	0.295	0.162
20–24	0.799	0.482	0.105	0.018	0.915	0.579	0.259	0.139
25–29	0.835	0.542	0.131	0.022	0.950	0.709	0.382	0.223
30–34	0.894	0.661	0.201	0.036	0.963	0.769	0.458	0.281
35–39	0.910	0.703	0.231	0.044	0.977	0.843	0.577	0.387
40–49	0.944	0:799	0.336	0.072	0.983	0.880	0.652	0.463
B Only mother illiterate (WI)								
15–19	0.843	0.555	0.138	0.024	0.905	0.550	0.237	0.126
20–24	0.784	0.460	0.097	0.016	0.889	0.505	0.206	0.107
25–29	0.822	0.520	0.121	0.021	0.934	0.643	0.314	0.175
30–34	0.884	0.641	0.185	0.033	0.950	0.711	0.385	0.224
35–39	0.903	0.683	0.216	0.040	0.969	0.801	0.505	0.319
40–49	0.939	0.784	0.316	0.065	0.977	0.844	0.582	0.389
C Mother primary (WP)								
15–19	0.843	0.557	0.138	0.024	0.894	0.517	0.214	0.112
20–24	0.784	0.460	0.099	0.017	0.875	0.473	0.187	0.095
25–29	0.823	0.520	0.121	0.021	0.926	0.613	0.289	0.157
30–34	0.885	0.643	0.185	0.033	0.944	0.684	0.354	0.203
35–39	0.903	0.686	0.217	0.041	0.965	0.777	0.473	0.293
40–49	0.939	0.784	0.319	0.065	0.974	0.828	0.550	0.361
D Mother primary + (W > P)[a]								
15–19	0.636	0.291	0.050	0.008				
20–24	0.540	0.217	0.034	0.005				
25–29	0.601	0.259	0.043	0.006				
30–34	0.713	0.368	0.069	0.011				
35–39	0.750	0.413	0.082	0.014				
40–49	0.835	0.540	0.130	0.022				
E Punjab								
15–19	0.821	0.518	0.121	0.021	0.913	0.573	0.254	0.136
20–24	0.757	0.422	0.084	0.014	0.898	0.528	0.221	0.116
25–29	0.798	0.481	0.106	0.017	0.940	0.664	0.335	0.189
30–34	0.868	0.604	0.164	0.029	0.955	0.730	0.407	0.241
35–39	0.888	0.649	0.191	0.034	0.972	0.814	0.526	0.339
40–49	0.930	0.757	0.284	0.057	0.979	0.856	0.604	0.412
F Sind								
15–19	0.834	0.541	0.131	0.022	0.907	0.553	0.239	0.126
20–24	0.773	0.444	0.092	0.016	0.890	0.506	0.208	0.108
25–29	0.812	0.503	0.114	0.019	0.934	0.644	0.317	0.177
30–34	0.878	0.626	0.176	0.031	0.951	0.714	0.388	0.227
35–39	0.896	0.669	0.205	0.038	0.970	0.801	0.506	0.321
40–49	0.935	0.773	0.302	0.062	0.978	0.846	0.582	0.392
G The NWFP								
15–19	0.835	0.543	0.132	0.022	0.960	0.755	0.439	0.266
20–24	0.775	0.446	0.093	0.016	0.953	0.720	0.395	0.232
25–29	0.814	0.506	0.115	0.020	0.973	0.820	0.536	0.348
30–34	0.879	0.630	0.179	0.032	0.980	0.861	0.614	0.421
35–39	0.898	0.671	0.207	0.038	0.988	0.910	0.720	0.541
40–49	0.937	0.775	0.304	0.063	0.991	0.932	0.778	0.616

[a]Category was excluded for rural subsample.



NOTE: Values of P(d) are estimated from the coefficients reported in table 12.
BI = both husband and wife illiterate
WI = wife only illiterate
WP = wife primary education
W > P = wife more than primary education (very few rural women have post-primary education).

Figure 17 Estimated proportions still being breastfed at ages 12, 18, 24 and 30 months, by age, education and urban-rural residence of the mother: logit-linear regression results for surviving children born in the 40 months preceding the survey (PFS) however, chosen model 4 for both urban and rural samples so that we could also consider the regional differentials.

The results obtained by fitting a logit-linear model can either be interpreted on the basis of the differences in the fitted values on the logit scale, or, more conveniently, by estimating the implied proportions. Table 12 presents by way of illustration the proportions of children still being breastfed at 12, 18, 24 and 30 months[56] since birth by selected characteristics of the mother, and figure 17 depicts the results for children born to mothers of ages 25–29 and 35–39 years. Once again the results are in the expected direction: at any duration since birth, children born to rural, illiterate and older mothers show a higher proportion still being breastfed. Educational effects become pronounced only when the mother has more than primary education, though we could examine this only for the urban sample. The results for urban-rural, age and educational differentials are generally in line with those obtained earlier (see figure 16). The only marked difference is an apparently greater rural-urban differential at long durations since the birth in the present analysis. Regional differences (not considered earlier) are not very marked in urban areas. In rural areas, however, the proportions are noticeably higher for children born to women of the NWFP than to those either of Sind or of the Punjab province. Differentials are greater among younger mothers than for the older mothers. For example, among children 12 months old, the estimated proportion still being breastfed was approximately five percentage points higher if the mother was from rural NWFP and aged 15–19 than if the mother was from rural Punjab and aged 15–19, whereas the proportion was only one percentage point higher for the rural NWFP mothers of 45–49, compared to the rural Punjabi women of the same age.

Proportional Hazards Life-Table Models

The application of the proportional hazards model to current status breastfeeding data is already well documented in McDonald (1981). In addition to describing the analytical aspects of the technique, McDonald explains both the methodological and practical problems (eg estimation procedures, use of different computer software, memory space requirements, and sample size considerations) in applying the model to breastfeeding data. The reader interested in actually applying the method should consult that paper. We shall restrict ourselves here to providing simply an illustration of the application of the model to the data we have already analysed using other techniques, in order to bring out the similarities and the differences.

The focus of the analysis is not the same as in the previous analyses. Those were based on the probability that a child would still be breastfed after a certain duration of time. We are now interested in the hazard function (which can be thought of as the force of weaning, taking the analogy of the force of mortality when events under study are deaths) or the instantaneous rate of weaning. In other words, we analyse the probability at time d of a child being weaned in the next instant of time given that he or she has not been weaned prior to time d. In so far as the socio-economic background characteristics influence these probabilities, they produce different hazard functions for children from different subgroups. In the proportional hazards model, the difference is modelled as a constant ratio of the hazard functions between subgroups across all values of d. Expressed another way, the hazards for any two individuals from different subgroups are proportional. The ratio of the hazards is sometimes called the relative risk; for example, the risk for a child born to an urban woman relative to the risk for a child born to a rural woman.

Taking the same independent variables as those used in the logit-linear model and retaining their categories (except for education for which all literate mothers were combined in a single category, irrespective of the level of education), we fitted several different models to the data sets. Again the urban and rural subsamples had to be analysed separately. The underlying assumption that the independent variables taken into account affect the hazard in a proportional manner was evaluated; the time (duration since birth) was subdivided and models were fitted to each category separately so as to satisfy the proportionality assumption more completely; interactions among the independent variables were considered; and the improvement in the fit of the model achieved by including a variable not previously in the model was tested. We have settled for the model with mother's age, education and region of residence (having additive effects among them) for both urban and rural subsamples.

The estimated relative risks for the various covariates are given in table 13. For age of mother we have taken 20–24 as the reference group, for education the lowest educational group, and for province Punjab. Once more the results are broadly comparable to those obtained with the other analyses, although some features emerge that were not

Table 13 Estimated relative risks of weaning for surviving children born in the 40 months preceding the survey, by age, residence and educational level of mother (PFS)

Background characteristics	Urban	Rural
Age		
15–19	0.829 (199)	0.838 (329)
20–24	1.000 (340)	1.000 (469)
25–29	0.860 (375)	0.711 (522)
30–34	0.667 (204)	0.566 (335)
35–39	0.611 (103)	0.438 (202)
40–49	0.419 (36)	0.362 (65)
Education		
Both illiterate	1.000 (431)	1.000 (1200)
Only wife illiterate	1.049 (493)	1.159 (619)
Wife literate	1.348 (333)	1.409 (103)
Province		
Punjab	1.000 (721)	1.000 (1336)
Sind	0.928 (464)	0.878 (363)
NWFP	0.877 (72)	0.584 (223)

NOTE: Figures in parentheses are the number of cases.

[56] For illustrative purposes, we have extracted the results for only the subgroups shown in figure 16. The proportions can, however, be calculated for any duration since birth category and for any combination of the categories of the socio-economic variables included in the model.

previously apparent (for example, the suggestion that the risk of weaning may be lower for children born to the very youngest mothers, aged 15–19, than for those born to women in their early 20s).

5.5 CONCLUDING REMARKS

Of the various techniques we have illustrated here, proportional hazards models and related methods constitute a particularly attractive set of tools for handling censored data sets for a wide variety of demographic phenomena that can be represented in the form of attrition tables. No doubt their use will become more and more common. However, there are both theoretical problems in the choice of the particular model to use, and considerable practical difficulties in their implementation. By contrast, the techniques illustrated first were extremely easy to apply.

The choice between the various possibilities will depend partly on the theoretical model considered appropriate, but it will also depend heavily on the facilities available. It is reassuring to see that for Pakistan, at least, the results are broadly similar for all four methods illustrated. The results we obtained using a simple exploratory analysis or using a simple regression analysis with polynomials are not markedly different from those obtained using the more sophisticated techniques.

6 The Impact of Breastfeeding on Fertility

6.1 UNDERLYING ISSUES

If estimation of breastfeeding patterns and differentials from WFS material is not entirely straightforward, even so it is probably less complex than evaluation of the impact of breastfeeding on fertility.

The most fundamental reason for this is the problem of reciprocal causality. Breastfeeding practices influence the time that will elapse before a women becomes susceptible again to conception, but the advent of a new pregnancy can, in its turn, have an influence on breastfeeding: many women wean their child completely only when they become pregnant again. Since respondents were not asked whether they stopped breastfeeding *because* they were pregnant, we cannot easily disentangle the two effects.[57]

A second set of issues has already been referred to in chapter 3. With what fertility variable should we relate breastfeeding in order to obtain the most meaningful analysis of the relationship between breastfeeding and fertility? With amenorrhoea? With the birth interval? Or with some other measure of fertility? None will provide an entirely pure picture of the relationship since each may also be affected by a number of other variables (Masnick 1979). Fortunately, in many populations the relationship between breastfeeding and fertility is so strong that its order of magnitude can readily be established even if we cannot fine-tune the picture to several decimal places. Moreover, introduction of controls for other variables may be not merely possible but also revealing.

A third set of reasons resides in technical difficulties stemming from the fact that we have only incomplete maternity histories for most women. Estimation problems arise in particular if we try to examine the relationship between duration of breastfeeding and length of the birth interval. These problems are discussed in section 6.3 below.

6.2 BREASTFEEDING AND AMENORRHOEA

Since the fertility impact of breastfeeding operates mainly through amenorrhoea, this relationship seems a logical starting point – perhaps even the best if we have to restrict ourselves to just one analysis. However, since the core questionnaire did not include questions on post-partum amenorrhoea, we cannot carry out a thorough analysis without the FOTCAF module. The only possibility open to us is the more questionable tactic of trying to predict the duration of amenorrhoea from the information on breastfeeding, by making the assumption that the average relationship found in other populations provides a good predictive tool for our population.

Several empirical relationships that describe the general relationship between average duration of amenorrhoea and duration of breastfeeding are available. A first group – Corsini (1979), for example – assumes a linear relationship, with each additional month in the mean duration of breastfeeding adding about half a month to the mean duration of amenorrhoea (and with mean amenorrhoea averaging 1.5–2 months in the absence of breastfeeding). A second group allows for the probable curvilinearity in the relationship with an additional month of breastfeeding having less impact on amenorrhoea at very long or very short durations of breastfeeding than at moderate ones. A curvilinear relation appears to provide a better fit. Of two curvilinear expressions available (Lesthaeghe and Page 1980; Bongaarts 1981), we have chosen to use that of Bongaarts here, primarily because it is more direct and simpler to use and because it is based on slightly more data points. Mean duration of amenorrhoea ($\bar{Y}$) is estimated from mean duration of breastfeeding ($\bar{X}$) as:

$$\bar{Y} = 1.753\, e^{(0.1396\,\bar{X}) - (0.001872\,\bar{X}^2)} \qquad (17)$$

With this relationship each additional month of breastfeeding in the range 0–6 or 30–36 months adds only about 0.3 months to the predicted mean duration of amenorrhoea, whereas in the range 6–15 or 25–30 months it adds about 0.6 months on average and in the range 15–25 months it adds about 0.8 months. Table 14 shows the resulting estimates of mean duration of amenorrhoea based on the mean durations of breastfeeding estimated in chapter 4 for all births in the four years preceding the survey.

If we assume that the average duration of amenorrhoea would be about two months in the absence of breastfeeding, then these calculations imply that in Pakistan breastfeeding adds 13 months on average to the period of post-partum non-susceptibility.[58] For births occurring to older women, to less educated and to rural women, the number of additional months of non-susceptibility is higher: 17 months for births to women 35–49 (compared with 12 months for those to women 15–24), 14 months for those to illiterate women married to illiterate men (compared with only 5 months for those to women with more than primary education), and 14 months for births to rural women (compared with 11 months for those to urban women).

Although these estimates may not be very precise, they do very clearly indicate that breastfeeding has a major impact on fertility *processes* in Pakistan, and that its impact varies significantly with the characteristics of the mother.

[57] To the extent that it is the duration of full rather than partial breastfeeding that determines the length of post-partum amenorrhoea (or abstinence), those countries that adopted the FOTCAF module and included the question on duration of full breastfeeding can use this variable in order to minimize the reverse causality. In most countries, though, this possibility is absent.

[58] This and the figures that follow refer to children who survived. Inclusion of children who died reduces the values by 1–3 months, as table 14 shows.

Table 14 Estimated mean duration of post-partum amenorrhoea following births in the four years preceding the survey (PFS)

	All ages	Age of woman at the birth		
		15–24	25–34	35–49
All children				
All women	12.8	11.5	13.3	15.9
Rural residents	13.8	12.2	14.3	16.6
Urban residents	10.4	9.5	10.7	13.0
Both husband and wife illiterate	13.6	12.0	14.1	16.7
Only wife illiterate	12.5	11.5	13.2	14.1
Wife primary	12.0	14.3	11.7	–
Wife > primary	6.6	6.2	7.3	–
Surviving children				
All women	15.1	13.9	15.3	18.6
Rural residents	16.1	14.8	16.4	19.1
Urban residents	13.0	11.5	12.3	16.6
Both husband and wife illiterate	16.3	14.7	16.4	19.5
Only wife illiterate	14.7	13.8	14.8	16.5
Wife primary	13.3	13.3	12.3	–
Wife > primary	7.4	6.8	9.7	–

NOTE: Duration of amenorrhoea was estimated simply as:

$$1.753e^{(0.1396\bar{X}) - (0.001872\ \bar{X}^2)}$$

where $\bar{X}$ is the estimated mean duration of breastfeeding (from tables 3, 5 and 6).

They do not tell us about the relationship between breastfeeding and fertility *levels* as such; the impact of breastfeeding might be offset by the effect of related variables. The remaining parts of this chapter consider two alternative ways of addressing this issue.

6.3 BREASTFEEDING AND BIRTH INTERVALS

The analysis of birth intervals often appears to be one of the most attractive and most direct ways of evaluating the relationship between breastfeeding and fertility. But a number of difficulties arise when we have only incomplete maternity histories for most women.

Imagine for a moment that we had complete maternity histories for everyone – that every woman had been followed to age 50, for example. We might well want to make a distinction between closed birth intervals on the one hand and each woman's final open birth interval on the other. Since the final open interval tends to be longer on average, the proportion of the interval that is attributable to lactational amenorrhoea (or to lactation-related abstinence or contraception) is likely to be lower for that interval than for the closed intervals.

When we have incomplete maternity histories, however, we are faced with the inescapable problem that we simply cannot tell which of the women are currently in a birth interval that will one day be closed and which are in an interval that will never be closed.[59] Any analysis that includes current open birth intervals will include an unknown and variable proportion of final open intervals where the role of breastfeeding is likely to be quantitatively different. The proportion will be small for young and recently married women (especially in societies where nearly every woman wants at least two children) and very large for women in their 40s, but we do not know more than that. This should not detract from the fact that any relationship we observe may be perfectly representative of all intervals: it merely means that we cannot make a distinction between two quantitatively different situations which we would want to make for a more refined analysis and that we must interpret the results with due caution.

Given these considerations, it may appear preferable to restrict analysis to intervals that have been closed before the survey. This strategy offers an additional advantage of practical simplicity. Since this data set includes only completed birth intervals (and completed durations of breastfeeding) with no censored cases, we can simply plot birth interval length against breastfeeding duration and use standard regression techniques – we do not have to use the special techniques for censored data. The price of computational simplicity, however, is selection biases and resulting difficulties of interpretation.

We have already seen that restriction of analysis to closed birth intervals leads to under-representation of long intervals, especially among women who have not been married very long. If breastfeeding and birth interval were perfectly correlated, then women with long breastfeeding would be equally under-represented, and there would be no problem. But the association is not perfect; in particular, some of the women with short birth intervals may have had relatively long breastfeeding. The data set may thus represent the full range of breastfeeding lengths fairly well, although it has selected out some women with longer than average birth intervals.[60] Figure 18 provides an illustration of the consequences of this for a hypothetical cohort of women in which breastfeeding bears the most commonly found positive relationship to interbirth interval. For simplicity, figure 18 illustrates the consequences for a subgroup of women who all start the given interval at the same moment in time. Each point indicates what would be observed for a given woman if she were followed long enough to close the interval; plus signs distinguish women who have not completed an interval at the time of the survey (predominantly those with long intervals) and who are therefore excluded from the last closed birth interval data set. The solid line represents the true relationship between breastfeeding and birth interval, the broken line the relationship that would be estimated from these data. Clearly, in this example, the data set would underestimate the impact of breastfeeding. Whereas in reality each extra month of breastfeeding corresponded with a ¾ month

[59] We can of course be fairly certain for those who are currently pregnant on the one hand and for those who have had a sterilization or report they have already reached menopause on the other, but for the majority of cases we usually cannot tell.

[60] More generally, a sample of last closed birth intervals is likely to be less representative of birth interval lengths than it is of the lengths of any of its components.

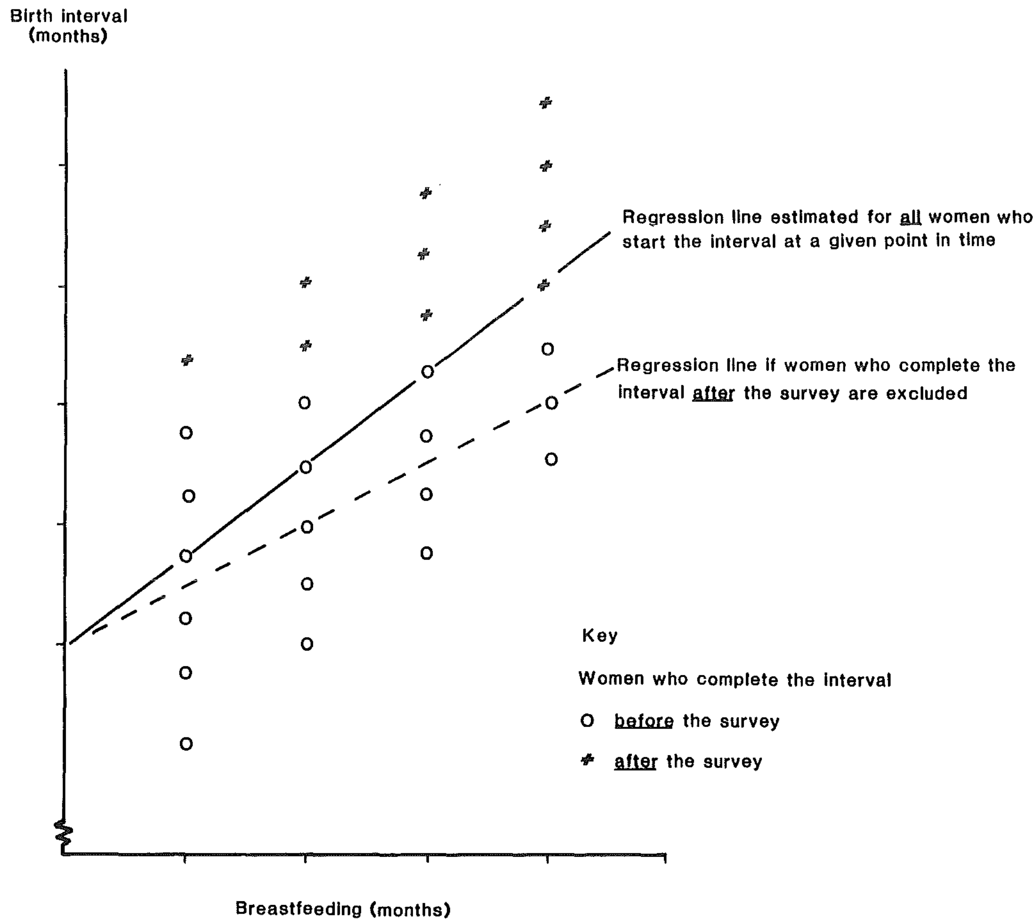


Figure 18 Apparent relationship between length of breastfeeding (or other components of birth interval) and length of the birth interval: illustration of the effect of restricting analysis to closed birth intervals

longer birth interval, the data set suggests that it corresponds with only a ½ month longer interval.[61]

To sum up, when we are working with incomplete maternity histories, there is no way to make a perfect analysis. We simply have to live with the problems, reducing the biases as much as possible and interpreting the results with care. Before proceeding to the substantive findings for Pakistan, we shall give a summary overview of the results obtained from several different ways of analysing the data. This overview is intended simply to bring out the relationship between the different possibilities so that the reader is in a better position to choose appropriate methods for her/his particular data set and to avoid overinterpretation of the results.

The overview is presented diagrammatically in figure 19. The analysis is age-specific (with age defined each time as age of the mother at the start of the birth interval in question), since factors determining other components of the birth interval (notably fecundability) and potential biases may vary significantly with age. Within each age group the intervals are divided into five categories according to the duration of breastfeeding in that interval.[62] For each age and breastfeeding subgroup, we have plotted bars showing the three quartiles – T_{25}, T_{50} and T_{75}, the times at which 25, 50 and 75 per cent respectively of the intervals have been closed. All the analyses are based on intervals for which we know how long breastfeeding lasted.[63] We can divide the analyses into three groups.

1 The first three analyses for each age and breastfeeding subgroup are all based on the data set defined by all intervals that were started in the four years immediately preceding the survey. In the examples illustrated here each interval has been given equal weight. Three variants are given:

F = 'Full data', ie questions on breastfeeding were posed for every birth in the period.

C = 'Core questionnaire' – a simulation of what would have happened if PFS had used the unmodified core questionnaire (breastfeeding data collected

[61] This particular example oversimplifies reality. In practice, since we are usually looking at intervals that started at different points in time, not all long intervals will be excluded (those that started a long time ago will not be excluded), and the bias illustrated here will, therefore, be muted although not eliminated.

[62] We have simply taken six-month wide categories here centred on preferred durations such as multiples of 6 and 12 months. In Pakistan very few women do not breastfeed or breastfeed for only a few weeks, and we have grouped these women together with the next category. The shortest category we have used is, therefore, 0–8 months.

[63] In other words we assume that for a given duration of breastfeeding, the probability of the survey being conducted before the child in question is weaned (ie intervals where we do not know how long breastfeeding will last) is not related to the length of the birth interval.

only for those births in the given period that were the two most recent births per woman)

M = 'Module' – a simulation as above but with the additional restriction embodied in the FOTCAF module that data were collected only for the one most recent birth for pregnant women.

All three analyses include current open birth intervals as well as closed intervals.

2 The second set of analyses also includes current open as well as closed birth intervals but is derived directly from the data on the two most recent births per woman regardless of when these occurred. The data differ thus from the first set in their time reference, which here is not the same for all women. In addition they provide more a 'per woman' analysis, whereas the first examples are a per interval approach. Two variants are illustrated:

C = 'Core Questionnaire' – breastfeeding questions posed for both the two most recent births for each woman.

M = 'Module' – breastfeeding questions restricted to just the most recent birth for pregnant women.

3 The final set of results illustrated is also a per woman approach with no fixed time reference, but is based on the last closed birth interval per woman, thereby excluding women who have had only one birth to date.

A glance at figure 19 shows immediately that, as expected for Pakistan, within a given age group longer durations of breastfeeding tend to be associated with significantly longer birth intervals. However, we can also see that within a given age and breastfeeding subgroup, the different analyses can yield markedly different estimates of the associated birth interval. Moreover, the way in which the birth interval increases with increasing duration of breastfeeding is not exactly the same for all types of analysis.

We compare first the two sets of analyses that include current open birth intervals as well as closed intervals, ie the per interval analyses based on all intervals started in the four years before the survey on the one hand (the first three bars in each panel) and the analyses based on the two most recent births per woman (the fourth and fifth bars). For the first two age-groups (intervals started when the mother was 15–24 or 25–34 years of age), as we might have expected intuitively, the per interval analysis yields shorter birth intervals for a given duration of breastfeeding than would be estimated from the two most recent births per woman: the first quartile (T_{25}), tends to be similar but the median (T_{50}) and, especially, the third quartile (T_{75}) tend to be lower. This is consistent with the fact that women with long intervals contribute a disproportionately low share of all births in a given period. We can see, however, that this pattern is *not* present for intervals started at advanced ages (35 years of age or older); except for intervals where breastfeeding was very short, the estimates based on all births in this age group are longer than those based on the two most recent births per woman. Clearly another factor is at work that becomes so important above age 35 that it compensates or even outweighs the per interval versus per woman differential. The reversal in the pattern results from the changed relationship between the two types of data in terms of the proportion of final open intervals each contains. For the oldest women, when we restrict analysis to intervals started in the last four years, a very high proportion of the intervals will be current open birth intervals, a high proportion of which again will be terminal open intervals. When we ask about the two most recent births regardless of when they occurred, more closed intervals will be included (even when we define age as age at the beginning of the interval). Since the terminal open interval tends to be longer than closed birth intervals, the higher proportion of terminal open intervals in the first data set pushes up the birth interval length relative to the second set.[64]

When we turn to the way in which birth intervals increase with longer breastfeeding, we find additional discrepancies. If we take all births in the four years before the survey, we find – as might be expected in a population with very little contraception – a monotonic rise in birth interval length as breastfeeding increases. If, however, we analyse the two most recent births we get a J-shaped relation – the birth interval is longer for intervals with very short breastfeeding (0–8 months) than for those with moderate breastfeeding (9–14 months) in both the central and the oldest age groups.

Within each type of analysis, there are differences depending on whether one has access to 'full' data, core questionnaire data or FOTCAF module data. The differences are quite marked at short durations of breastfeeding: for intervals started in the last four years where breastfeeding was only 0–8 months, the module (third bar) would have overestimated median birth intervals (T_{50}) by 5–15 months! At longer durations the differences are smaller, almost disappearing at breastfeeding durations above 15–20 months.[65] Clearly if the PFS had used the unmodified core questionnaire, and more so if it had used the FOTCAF module, we would have both rather seriously overestimated the length of the birth interval in the absence of breastfeeding and underestimated the rate at which the birth interval increased as breastfeeding increased. Expressed another way, the core questionnaire and even more so the FOTCAF module tend to yield conservative estimates of the relation between breastfeeding and birth interval. Similarly, if we had based our analysis on the two most recent births for each woman regardless of when they occurred, the restrictions imposed in the FOTCAF module would have led to apparently longer birth intervals at very short durations of breastfeeding; again the module would have underestimated the relationship between breastfeeding and birth interval.

Finally, we can turn to the analyses based on the last closed birth interval for each woman (the sixth bar in each panel). Since these are per woman estimates, the logical first comparison is with the fourth and fifth bars – the estimates based on the two most recent births per woman. With the last closed birth interval, both the restriction to

[64] We can note that the application of differential weights to the births in the last four years to give equal weight to each woman who had given birth in this period would reduce the difference but would not entirely eliminate it.

[65] In Pakistan, as in most countries, short breastfeeding is associated with short birth intervals. Since it is short birth intervals that are under-represented in the core questionnaire relative to full data and in the module relative to the core questionnaire, the impact of the restrictions is greatest at short durations of breastfeeding.

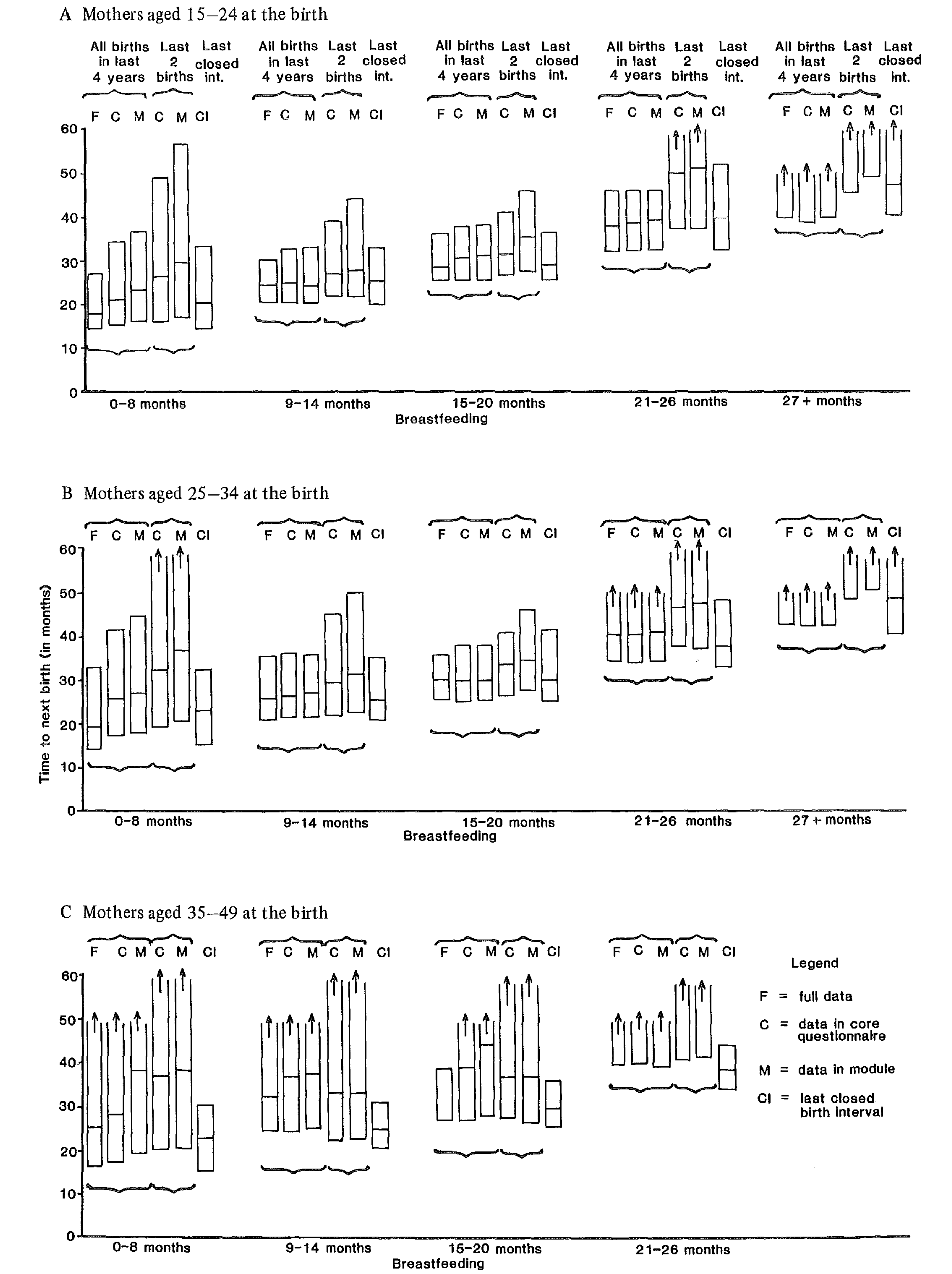


Figure 19 Median (T_{50}) and quartile (T_{25}, T_{75}) time to closure of birth interval by duration of breastfeeding (PFS)

women with two children (which leads to under-representation of long interbirth intervals, particularly at young ages) and the exclusion of open intervals (which leads to the exclusion of those final open intervals which are more common at older ages) pull down the estimated birth interval lengths and dramatically reduce their spread. In fact, the PFS estimates based on the last closed birth interval per woman are much closer to the estimates based on all intervals started in the four years before the survey than they are to those based on the two most recent births per woman. At younger ages, the downward selection bias almost exactly balances out the per interval versus per woman differential. Since we have not made this analysis birth-order specific, the selection bias in the last closed interval should be negligible for the oldest age group (virtually every woman has had ample time to have at least two children). We can notice in passing, however, that for this age group the estimated birth intervals derived from the last closed birth interval are shorter (rather than longer) than the estimates derived from all births in the four years before the survey. This reflects the extent to which the estimates based on all intervals rather than on closed intervals are pushed up at older ages by the inclusion of a significant proportion of terminal open intervals.

What can we conclude from all this? First, age-specification is obviously essential. Secondly, where fertility is high and contraception rare, we suggest that the relationship be analysed two ways, concentrating on all intervals started in the last four years (either with equal weight given to each interval or to each woman), on the one hand, and on the last closed birth interval for each woman on the other. The latter is likely to underestimate birth interval lengths for women at short durations of marriage and to give conservative estimates of the relationship between breastfeeding and birth interval for these women. The former is likely to be particularly hard to interpret at older ages where the proportion of current open intervals that are terminal open intervals starts to increase rapidly. In addition, we should repeat the caveat for all births in the last four years that if the data come from the unmodified core questionnaire, and in particular, if they come from the FOTCAF module, they too are likely to underestimate the relationship. Thirdly, where fertility is low, or among contraceptors, it may be more reasonable to concentrate solely on the last closed interval.

Turning now to the substantive findings, figure 20 plots the proportion of intervals that have not been closed by time elapsed since the interval started. The distributions are

A Estimates based on last closed birth interval per woman

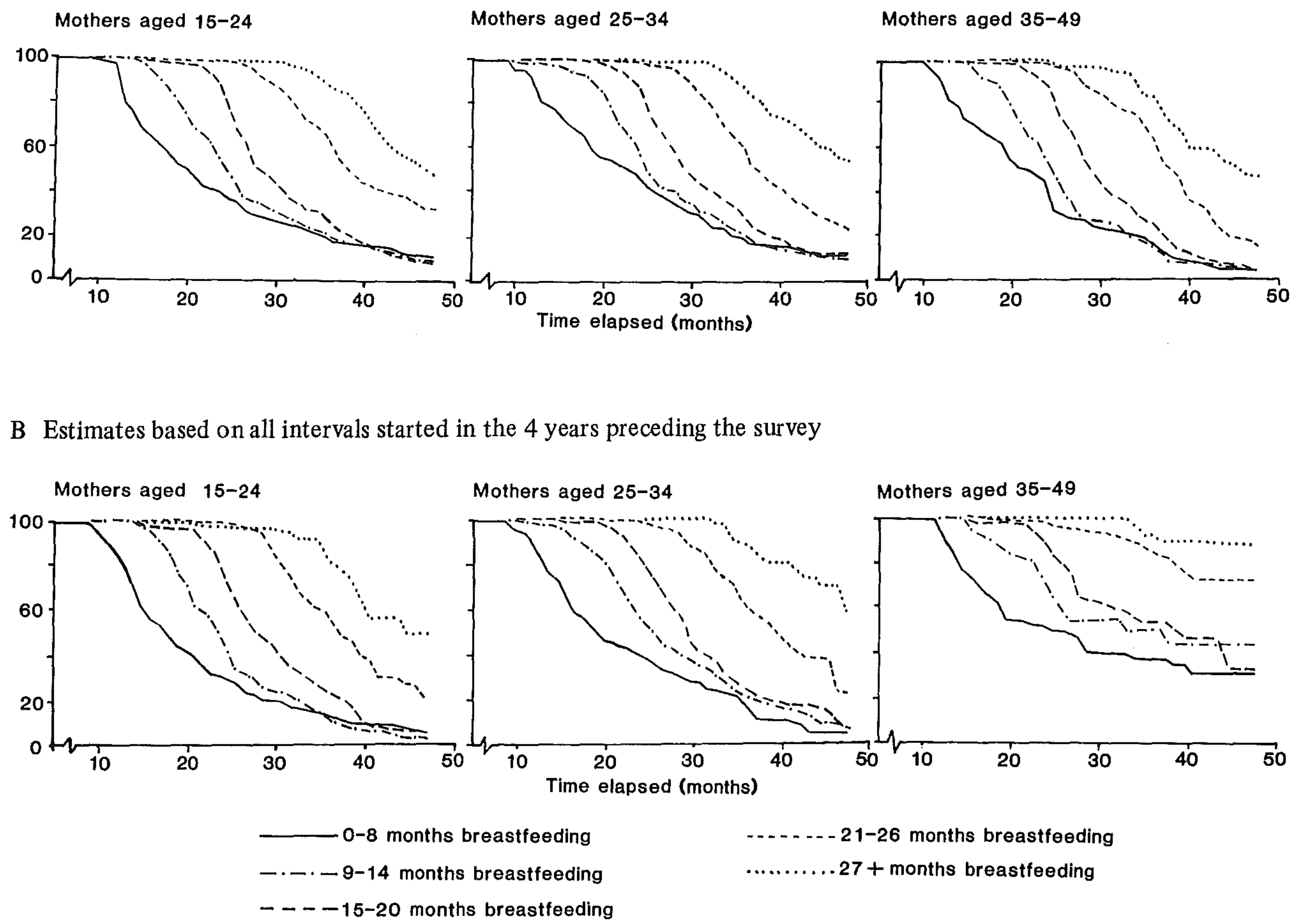


Figure 20 Proportion of birth intervals not yet closed, by time elapsed since start of the interval and duration of breastfeeding in that interval, by age of mother at start of the interval (PFS)

totally different for the five different categories of breastfeeding duration. The time elapsed before a given proportion of intervals is closed is 20–25 months longer for the longest breastfeeding category than for the shortest. There is a fairly systematic difference of about 5 months between adjacent categories, ie categories centred on durations that differ by about 6 months. As an average this is equal to, or somewhat higher than, one might perhaps expect on the basis of the probable impact of breastfeeding on amenorrhoea. We can also note two systematic deviations from the regular way in which 1 month more breastfeeding corresponds with almost 1 month more in the birth interval on average. The difference between the third and fourth categories of breastfeeding (15–20 months and 21–26 months respectively) seems to be larger, and the differences between the first three categories diminish with time and have almost entirely disappeared within three years of the start of the interval.

For a given duration of breastfeeding, the practice of contraception could of course extend the birth interval; conversely the death of a child might result in a tendency to close an interval that would have otherwise remained open (replacement effect) or to close it more quickly than if the child had lived (different child-spacing needs). We have run separate analyses to investigate these possible effects in Pakistan; the results are summarized in table 15 in the form of the T_{50} values. Not very surprisingly, given the very low use of contraception in Pakistan, the results for intervals to never-users of contraception are essentially similar to those for the whole sample. The results for intervals where the starting child died are distinct, however, and even somewhat surprising at first sight. Whether we analyse the last closed birth interval data or the data for all intervals started in the last four years, *less* time elapses before 50 per cent of the intervals are closed when we restrict analyses to intervals where the starting child survived than when we include intervals where the child died. In other words, for a given duration of breastfeeding, the next birth comes later, not sooner, if the child dies.

This pattern may at first appear surprising, but there are several reasons why we might expect to find it. First, the hypothesis that the next birth tends to be realized sooner if the child dies is based partly on the notion that breastfeeding will tend to be shorter in cases where the child died, and here we have controlled for duration of breastfeeding. Secondly, it is full breastfeeding rather than partial breastfeeding that influences the post-partum non-susceptible period. Where breastfeeding is truncated by the death of the child, the proportion of the total period of breastfeeding that is partial breastfeeding is lower and the proportion that is full breastfeeding is higher. In other words, for a given total duration of breastfeeding, the duration of full breastfeeding (and hence, on average, the duration of post-partum non-susceptibility) is likely to be longer where the child dies than where it survives, and this tends to push the overall birth interval up. Related to this is the possibility that where the child survives the woman has a greater chance of continuing breastfeeding until she becomes pregnant again. If she weans her child only when she is pregnant again, the birth interval will contain no segment representing the waiting time between weaning and the start of the next gestation; on the contrary, the periods of breastfeeding and gestation will overlap. In other words the birth interval will be (breastfeeding duration + 9 months' gestation – any overlap) rather than (breastfeeding duration + waiting time + 9 months' gestation). That breastfeeding is quite frequently continued up to or even into the next pregnancy in Pakistan is confirmed elsewhere in the PFS, where nearly 10 per cent of the women reported as currently pregnant at the survey (62 out of 654) were still breastfeeding.

Unfortunately it is not always a straightforward matter to separate out all the instances where pregnancy preceded weaning and where the birth interval was thus more a determinate than a consequence of the duration of breastfeeding. Operationally, it would be a simple matter to exclude those intervals that were shorter than the stated duration of breastfeeding plus a nine-month allowance for gestation. Where data are of very high quality, the results could be extremely interesting; but where a significant proportion of birth dates are imputed or where there is a sizeable margin of error in the individual reports of breastfeeding duration (due to rounding, for example) the picture could be quite misleading. Imputation or misreporting could go either way; the apparent birth interval might be too long relative to the reported breastfeeding or it might be too short. Exclusion of intervals where the next conception apparently preceded weaning would eliminate cases where imputation or misreporting had created a birth interval that was short relative to its breastfeeding, while retaining those cases where imputation or misreporting had created a birth interval that was long relative to its breastfeeding. The exclusions would introduce a systematic bias towards cases where the apparent birth interval was on the long side, given its reported breastfeeding.

The analyses discussed so far have all been based on grouped data using categories of breastfeeding duration. For the last closed birth interval, we can also carry out standard regression analysis using individual-level data with both breastfeeding and birth interval reported in single months. A simple linear regression model confirms the general picture already sketched: the birth interval is of the order of 19 or more months in the absence of breastfeeding, depending on age, and increases by about one month for every additional month of breastfeeding. It also draws attention to another feature that we have not mentioned explicitly before: the r^2 values are fairly low (about 0.20). A visual inspection of the scatter plots confirms immediately that the low values do not result from choice of an inappropriate specification of the relationship but rather from the huge scatter of the points about *any* line one might try to fit, straight or otherwise. For a given number of months of breastfeeding, there is a large variation in the duration of the associated interval, and vice versa. Despite the strong association between the two variables at the aggregate level, the association is weak at the level of individual intervals.

6.4 BREASTFEEDING AND CUMULATED MARITAL FERTILITY

An alternative approach for studying the relationship between breastfeeding and fertility is based on cumulated fertility (parity) rather than on just one or more specific birth intervals. Since parity embodies all the components of fertility that might reinforce or compensate breastfeed-

Table 15 Median duration of birth interval by duration of breastfeeding, for all intervals started in the four years preceding the survey and for the last closed birth interval per woman (PFS)

Age of mother at start of interval	Duration of breastfeeding (in months)				
	0–8	9–14	15–20	21–26	27+
	Median birth interval[a] (in months) Intervals started in the 4 years before the survey				
15–24					
All intervals	17.5	23.5	28.1	37.0	–
Non-contracepting mothers	18.0	23.3	29.5	37.5	–
Intervals where 1st child survived	16.3	23.4	27.7	38.0	–
Intervals where 1st child survived, for non-contracepting mothers	16.6	23.2	28.3	37.3	–
25–34					
All intervals	18.8	25.1	29.1	39.3	(⩾48.0)
Non-contracepting mothers	18.9	24.4	28.6	39.1	(⩾48.0)
Intervals where 1st child survived	18.0	23.9	28.9	40.7	(⩾48.0)
Intervals where 1st child survived, for non-contracepting mothers	(17.7)	24.0	28.8	39.8	(⩾48.0)
35–49					
All intervals	(24.5)	(32.5)	(38.5)	(⩾48.0)	–
Non-contracepting mothers	(26.9)	(26.2)	(38.9)	(⩾48.0)	–
Intervals where 1st child survived	–	(25.1)	(37.6)	(⩾48.0)	–
Intervals where 1st child survived, for non-contracepting mothers	–	–	–	(⩾48.0)	–
	Median birth interval[a] (in months) Last closed birth interval per woman				
15–24					
All intervals	19.4	24.3	28.2	38.2	46.9
Non-contracepting mothers	19.8	24.6	28.3	38.7	47.5
Intervals where 1st child survived	18.1	24.2	28.5	38.1	47.3
Intervals where 1st child survived, for non-contracepting mothers	(18.2)	24.3	28.6	39.4	48.3
25–34					
All intervals	23.0	24.7	29.2	37.2	47.9
Non-contracepting mothers	23.5	25.0	28.9	37.6	47.2
Intervals where 1st child survived	19.4	25.3	28.5	37.2	47.8
Intervals where 1st child survived, for non-contracepting mothers	18.7	24.3	28.5	37.8	47.6
35–49					
All intervals	(22.3)	(24.4)	29.0	(38.0)	–
Non-contracepting mothers	(23.6)	(25.1)	(30.0)	(38.3)	–
Intervals where 1st child survived	–	(24.3)	29.1	(36.5)	–
Intervals where 1st child survived, for non-contracepting mothers	–	(24.5)	(30.8)	(38.3)	–

[a]Median birth interval defined as time at which 50 per cent of the intervals concerned have been closed.
NOTE: ( ) = N < 100
– = N < 50

ing differentials, including those that can occur at quite different points in a woman's life, this approach is particularly suited to bringing out the overall relationship between breastfeeding and fertility.

An appropriate measure of cumulated fertility is provided by the DRAT-index proposed by Boulier and Rosenzweig (1978), which can be calculated for individuals. Before proceeding to analysis of the observed relationship between breastfeeding and DRAT values, however, a word has to be said about DRAT itself.

DRAT is simply a standardized index of a woman's cumulated fertility to date, standardized for net exposure to reproduction; the exposure period is defined as the age range over which she has been in a marital union. The value of DRAT for each woman is simply the ratio of her observed parity to the average parity expected for women with her age at marriage and current age if they had been subject in the intervening period to a standard schedule of age-specific marital fertility rates n(x). For a currently married woman who married at age m and who is aged a at the time of the survey, the expected parity[66] is

$$\int_m^a n(x)\,dx$$

and her DRAT value is defined as

$$C(a, m)/\int_m^a n(x)\,dx \qquad (18)$$

where C (a, m) is her observed parity. A value of DRAT of 0.75, for example, means that the woman has achieved only three-quarters of the fertility she would have had if subject to the standard schedule.[67]

The standard schedule most often used is that offered by Coale, Hill and Trussell (1975), which reflects the age pattern of fertility common to several natural fertility populations with accurate data. Since records from historical European populations predominate in the source data set they used, the age pattern of fertility embodied in the standard essentially reflects historical European experience. For analysis of differentials within a country, one may sometimes prefer to use a local standard, based for example on the national age-specific marital fertility rates.

There are three major assumptions underlying the use of DRAT values regardless of which standard schedule is used:

1 current parity should be reported accurately for all women, irrespective of their ages for example;
2 the shape of the standard age schedule of fertility, n(x), should be appropriate; this implies that
3 the simplification introduced by letting the standard schedule depend only on age should hold.

There are reasons to believe that none of these three assumptions are met completely and that the resulting distortions may be serious enough to require appropriate corrective measures.

First, observed DRAT values are likely to decline artificially with age because of under-reporting of parity by older women. Any multivariate analysis involving parity data should, therefore, incorporate a control for age of respondent in order to eliminate the potential impact of this artefact on the differentials estimated for other variables.[68] The second and third assumptions are closely interconnected; natural fertility rates do not always depend on age alone but can also vary, at a given age, with duration of marriage. The marital fertility rates in a particular age group, say 30–34, are typically lower for women who have been married a long time than for those who have entered marriage more recently.[69] In a sense, a marriage 'ages': women who marry before age 15 for example will have had 20 years of potential childbearing by the time they reach age 35 and their marital fertility rate beyond age 30 may be substantially lower than that of women who started reproduction much later. By age 35, many early marriers have had time to achieve a sizeable family and to have reached a parity they do not want to exceed; they may also have adult children or already be grandmothers, and in many cultures actual or potential grandparental status is considered incompatible with further childbearing. There may also be non-volitional fecundity impairments that increase with marriage duration or parity. The effect of age at marriage on the age pattern of marital fertility rates may be relatively insignificant in populations where marriage is very late, but in those where many women marry early it may well be significant.[70] Whatever standard we use, it embodies a particular age pattern of marriage and it is likely to decline too rapidly with age for populations or for subgroups characterized by later marriage and too slowly for those characterized by earlier marriage. For the first set (groups with later marriage), their estimated DRAT value will be artificially inflated above about age 30 and for the second set (characterized by earlier marriage), the DRAT values will be artificially reduced beyond about age 30.[71] Multivariate analyses need to incorporate, therefore, not only the control for age introduced above but also a control for age at marriage.

In the analysis that follows we try to evaluate the contribution of breastfeeding durations to DRAT values relative to that of other socio-economic variables (ethnic or linguistic group, education, exposure to mass media and use of contraception). A preliminary plot of the mean DRAT values by age and by age at marriage was made to see whether the relationships were sufficiently close to linear for us to introduce a control for the effects of age and age at marriage simply by introducing these two variables as covariates in a multiple classification analysis, intended to examine the impact of the other variables. The plots were close to linear, so this procedure was adopted. We should note, however, that to keep the number of possible variable combinations within the bounds of most computer packages, other variables must often be introduced as broad categories, ie as factors (even if they are really continuous variables), and possible interaction effects are ignored.

Two restrictions have to be put on the analysis. It is restricted to women who have been married at least five years and to women with at least one child who has been weaned.

---

[66] Assuming that she has remained married all the time.

[67] If we can assume that the age pattern of the standard schedule is appropriate for our population, then a synthetic total marital fertility rate can be calculated by multiplying DRAT by the TMFR of the standard schedule (9.1 from age 20 onward, and 11.8 from age 15 onward in the standard schedule used here).

[68] It might of course be possible that the decline in DRAT values with age reflects a real rise in fertility among younger women, but other checks have to be used to establish whether this alternative hypothesis is plausible or not.

[69] We can note that the opposite may occasionally occur when contraception is quite widespread, namely those who marry earlier sometimes have systematically higher fertility at all ages than those who marry later, because they are a selected group with less effective contraceptive use.

[70] The Pakistan data illustrate the problems rather well. Currently married women above age 45 who had never used contraception reported a mean age at last birth of only 35.5 years compared with the values of 40–42 years typically found in historical European populations. Moreover, women over 45 whose first born child was currently 30 years or over reported a median age of 31 years at the birth of their last child, whereas those whose first born child was currently 20 years or less reported a median age of 41 years.

[71] They would also be reduced by any periods of widowhood, divorce or separation that were not allowed for.

The first restriction is necessary because DRAT is a rather volatile index during the first few years of marriage. The volatility derives both from the nature of fertility and from the nature of our index. With respect to the first, we know that chance factors (such as waiting time to conception) have a relatively large impact on achieved parity during the early years of marriage; it is only after several years that systematic factors stand out more clearly over the chance ones. With respect to the second, our index is composed of a denominator that is very small in the early years of marriage ('expected' parities of a fraction of a child) combined with a numerator that increases by leaps (from actual parity 0 to actual parity 1, 2, etc). Both these characteristics can lead to large sampling variations in average DRAT values, especially for small subgroups. Hence the recommendation that women married less than five years be excluded from the analysis.

The second restriction is introduced in order to have information on at least one completed duration of breastfeeding for each woman. Ideally, since DRAT is a measure of lifetime fertility experience up to the date of survey, the other variables should also measure cumulated experience. For breastfeeding, for example, we should take for each woman her average duration of breastfeeding measured over all children she has weaned. Since most WFS surveys do not have breastfeeding data for all births, we are forced in most countries to use the data from the most recent births as a proxy. We have analysed the data for Pakistan both ways, in order to illustrate the impact of this restriction.

In the analyses presented here (tables 16 and 17), the five factors included were introduced in a stepwise fashion that largely reflects their temporal sequence: ethnic or linguistic group was introduced first, followed by education and (current) exposure to mass media, while use of contraception and duration of breastfeeding were introduced in the final step. The Coale–Hill–Trussell standard schedule was used here, which gives an overall mean DRAT value of 0.74. The extent to which each sub group differs from this mean is shown in column 1 for the zero-order deviations and in columns 2–6 for the deviations that emerge at each of the steps introducing other variables. At each step the deviations were adjusted for the two covariates – age and age at marriage – as well as for the other factors present in the regression at that step. The increment in R^2 that occurs as successive variables are introduced can thus also be studied (bottom line).

Looking first at table 16 which is based on the preferred definition of breastfeeding experience (the average duration to date for each woman), we see that of all the variables introduced in the analysis, the duration of breastfeeding stands out as the variable associated with by far the largest variation in DRAT values. This is true whether we look at the unadjusted deviations for each variable (column 1) or at the deviations adjusted for the effects of the other variables (columns 2–6). From column 1 in table 16, for example, we see that women in the shortest breastfeeding category used (0–8 months)[72] had a mean DRAT value as much as 16 units (= 21 per cent) higher than the national average, whereas those in the longest breastfeeding category (⩾ 27 months) had a DRAT value as low as 34 units (= 45 per cent) below the average. The DRAT-value of these short breastfeeders was more than twice as high as that of the long breastfeeders. These deviations are only very slightly reduced (from 16 to 14 and from 34 to 33) even when all the other factors and the two covariates are included (column 6). The increment in R^2 that occurs when breastfeeding is introduced ($R^2 = 0.275$ compared with $R^2 = 0.076$ before introduction of breastfeeding) can also be seen to be overwhelmingly larger than the increment that could be contributed by any of the other four factors or the two covariates. Clearly breastfeeding behaviour is the dominant variable considered. This may seem surprising in the light of classic fertility transition theories, from which one might expect the education factor to emerge more prominently. The results may also be surprising in the light of widely held views of the mechanisms of fertility transition, in which contraceptive use is seen as the major means by which modernization in general, and education in particular, reduce marital fertility. However, both education and contraceptive practice exhibit deviations that are extremely revealing, if at first sight perhaps surprising.

Use of contraception is associated with positive deviations above the average DRAT value rather than with negative ones, and this for both the unadjusted and the adjusted deviations. In other words, both in the population in general and within the various socio-economic subgroups considered, there is a tendency for the women with the highest DRAT values to be those who report they have used contraception. The data suggest a significant selection process in the adoption of contraception; it appears to be more a situation in which higher than average fertility to date provokes adoption of contraception than one in which contraception is adopted in order to achieve lower than average fertility.

For education, the unadjusted deviations in DRAT exhibit almost no differences at all between educational levels. Of all the variables included in this analysis, education has the lowest eta value (0.1).

Two sets of factors may underlie these patterns. First, our analysis covers all women married at least five years, including those still in relatively early stages of their family formation. Even if their family size preferences are lower than traditional preferences, these women may not yet have reached the point at which they will call a halt to further childbearing. Secondly, if we examine the adjusted deviations more closely, we see that these suggest that the apparent lack of an overall education effect may be the result of significant trade-offs. When adjusted for contraceptive use and for breastfeeding duration, the DRAT deviations show slightly more variation between education categories (beta = 0.10) and, moreover, this time the deviations are in the expected direction with higher educational levels being associated with slightly lower

[72] In populations where a sizeable proportion of women never breastfeed (or breastfeed for only extremely short durations), it may be desirable to create a separate category for them. Unless these women are thought to be characterized by very different fecundity or fertility behaviour, however, this is probably not necessary. The differences in mean duration of amenorrhoea are small at short durations of breastfeeding (using equation (17), the difference in predicted amenorrhoea between the two limits of our shortest category – no breastfeeding at all and 8 months – would be only three months). Interestingly enough, creation of a separate category 0–2 months in Pakistan produces a group with a negative rather than a positive deviation in DRAT (–0.06 unadjusted, –0.09 adjusted, but this is based on a rather small number of women (n = 124)).

Table 16 Multiple classification analysis (MCA) of the relative level of cumulated marital fertility (DRAT), incorporating breastfeeding data for all births: all women married five years or more with at least one weaned child (PFS)

Variable and category	N	Unadjusted deviation from the overall mean[a] (1)	Overall mean DRAT: 0.75 (N = 3727) Deviations from the overall mean adjusted for other factors and covariates[b] (2)	(3)	(4)	(5)	(6)
Ethnic group							
Urdu	325	0.04	0.05	0.06	0.05	0.04	0.01
Punjabi	2630	–0.01	–0.01	–0.01	–0.01	–0.01	0.00
Pushto	241	0.06	0.06	0.06	0.06	0.06	0.08
Sindhi	531	–0.02	–0.02	–0.03	–0.03	–0.02	–0.05
		(0.08)	(0.08)	(0.08)	(0.08)	(0.07)	(0.10)
Index of education							
Both illiterate	2251	–0.00		0.01	0.01	0.01	0.01
Wife illiterate, Husb. primary	583	0.01		0.00	0.00	0.01	0.01
Wife illiterate, Husb. secondary +	551	0.01		–0.01	–0.01	–0.02	–0.02
Wife primary	222	0.00		–0.02	–0.03	–0.04	–0.04
Wife secondary +	120	–0.01		–0.06	–0.07	–0.11	–0.12
		(0.02)		(0.05)	(0.06)	(0.09)	(0.09)
Exposure to mass media							
Not exposed	2465	–0.01			–0.01	–0.01	–0.00
Exposed	1262	0.02			0.02	0.01	0.00
		(0.05)			(0.05)	(0.02)	(0.00)
Use of contraceptives							
Never used	3260	–0.01				–0.02	–0.01
Ever used	467	0.09				0.11	0.08
		(0.12)				(0.14)	(0.10)
Average length of breastfeeding (months)							
0–8	505	0.16					0.14
9–14	1222	0.11					0.10
15–20	1167	–0.02					–0.01
21–26	657	–0.19					–0.19
27+	176	–0.34					–0.33
		(0.48)					(0.46)
R^2			0.055	0.057	0.058	0.076	0.275
Partial R^2 (due to factor added)				0.002	0.001	0.019	0.215

[a]Values in parentheses are eta or beta-coefficients.
[b]Covariates are age (adjusted b = –0.006, $p < 0.001$) and age at marriage (adjusted b = 0.004, $p < 0.001$).

DRAT values. It is probably largely because it is precisely the women with the highest educational exposure who tend to have the shortest breastfeeding that their overall (unadjusted) DRAT value is not below average.

Overall, although this particular analysis is too limited to be conclusive, it is consistent with what may be a widespread pattern of fertility change in the earliest stages of transition. At a very early stage the traditional restraints on marital fertility, particularly those that operate through child-spacing, may decline; the potential for fertility increase that is thereby released may not immediately be compensated by new restraints such as contraception, abortion and sterilization. In a few cases, fairly complete compensation may occur if contraception, etc are adopted as alternative means of spacing, although this form of compensation often seems to be less than complete (Page and Lesthaeghe, eds, 1981). Ultimately the affected cohorts are likely to reach their desired family size earlier than more traditional women and may well adopt contraception and other forms of fertility restriction at that point. Moreover, since it is usually the most modernized group who reduce their traditional child-spacing patterns first and it is this subgroup who often express the smallest preferred family sizes, their final fertility may well in the end be lower. In

Table 17 Multiple classification analysis (MCA) of the relative level of cumulated marital fertility (DRAT) – effect of restricting breastfeeding data to the two most recent births: all women married five years or more with at least one weaned child (PFS)

Variable and category	Unadjusted deviations from the overall mean[a]			Overall mean DRAT = 0.75 Deviations from the overall mean adjusted for other factors and covariates[b]		
	Breastfeeding data			Breastfeeding data		
	All births (1)	2 most recent births (2)	1 birth per woman (3)	All births (4)	2 most recent births (5)	1 birth per woman (6)
Ethnic group						
Urdu	———	0.04	———	0.01	0.02	0.03
Punjabi	———	–0.01	———	0.00	–0.00	–0.00
Pushto	———	0.06	———	0.08	0.07	0.07
Sindhi	———	–0.02	———	–0.05	–0.05	–0.04
	———	(0.08)	———	(0.10)	(0.09)	(0.08)
Index of education						
Both illit.	———	–0.00	———	0.01	0.01	0.01
Wife illit., Husb. prim.	———	0.01	———	0.01	0.01	0.01
Wife illit., Husb.sec. +	———	0.01	———	–0.02	–0.02	–0.02
Wife primary	———	0.00	———	–0.04	–0.04	–0.04
Wife sec. +	———	–0.01	———	–0.12	–0.12	–0.11
	———	(0.02)	———	(0.09)	(0.10)	(0.09)
Exposure to mass media						
Not exposed	———	–0.01	———	–0.00	–0.00	–0.00
Exposed	———	0.02	———	0.00	0.00	0.01
	———	(0.05)	———	(0.00)	(0.01)	(0.02)
Use of contraceptives						
Never used	———	–0.01	———	–0.01	–0.01	–0.01
Ever used	———	0.09	———	0.08	0.09	0.10
	———	(0.12)	———	(0.10)	(0.12)	(0.12)
Length of breastfeeding (months)						
0–8	0.16	0.12	0.05	0.14	0.11	0.04
9–14	0.11	0.07	0.09	0.10	0.06	0.08
15–20	–0.02	0.02	0.03	–0.01	0.02	0.03
21–26	–0.19	–0.08	–0.08	–0.19	–0.08	–0.07
27+	–0.34	–0.17	–0.18	–0.33	–0.15	–0.16
	(0.48)	(0.32)	(0.28)	(0.46)	(0.29)	(0.26)
R^2				0.275	0.152	0.129
Partial R^2 (due to addition of breastfeeding)				0.215	0.092	0.067

[a]Values in parentheses are eta or beta coefficients.
[b]Covariates are age (adjusted b = 0.006 to 0.007, $p < 0.001$) and age at marriage (adjusted b = 0.004 to 0.006, $p < 0.001$).

the meantime, however, their pace of family formation may be no different or may even be faster than that of more traditional groups. Such a pattern may be developing in Pakistan, although if so it is still at a very early stage. We should note, however, that it could be occurring faster than these particular results suggest because of the way we have treated age in this particular analysis. We have not examined age differentials as such (age having been introduced as a covariate partly for other reasons), whereas we know from the preceding chapter that children born to younger women are breastfed several months less than those born to older women.

Finally, we return to the question that is very relevant for other WFS data sets. What would have happened to our analysis if the PFS had not included questions about breastfeeding for all the births and we had estimated each woman's breastfeeding experience from data on the two most recent births? An illustration is provided in table 17, where we have compared the unadjusted and the fully adjusted deviations obtained in the preceding analyses (reproduced here for convenience in columns 1 and 4) with the results obtained using just the two most recent births.[73] In columns 2 and 5 we have made maximum use of the data

[73] Results for the intermediate steps in which breastfeeding was not introduced are not given here as they remain the same whichever definition of breastfeeding is used.

on the two most recent births. For women who had weaned only one child, we took the breastfeeding duration for that child, but for those who had weaned both of the two most recent children, we took the average of the two durations. This is the best we can do with a standard core questionnaire and somewhat better than we could do with the FOTCAF module. Columns 3 and 6 illustrate what would happen if we had restricted ourselves to data for just one birth per woman.[74]

The apparent impact of breastfeeding is markedly reduced if only one or two births are considered. In this example the deviations, the eta and beta values and the increment in R^2 decline to only half their true values if only one birth per woman is taken into account, and to only two-thirds of their true value even if use is made of information for both of the two most recent births. The attenuation effect arises simply from the fact that a woman's experience at a particular birth may be quite untypical of her overall experience, and the fertility-impact of just one breastfeeding experience can be attentuated by the impact of her breastfeeding behaviour at other births. Expressed another way, when cumulated experience is considered, only women with systematically very short or very long breastfeeding fall in the extreme categories, but when only one birth is considered, these categories include some women whose average or cumulated breastfeeding experience (and hence their cumulated fertility) is more moderate.[75]

The lesson to be drawn from this illustration for analysis of either the standard core questionnaire or the FOTCAF module is clear. Analysis of the relationship between breastfeeding behaviour and cumulated fertility can quite seriously underestimate the impact of breastfeeding if the breastfeeding data are restricted to just one or two births per woman. In Pakistan the true coefficients for breastfeeding were so much larger than those for any of the other variables considered that even when they were seriously underestimated (halved!) they remained the largest. This will not necessarily occur in other countries.

---

[74] In this particular example we took breastfeeding in the last closed birth interval for women with two or more children, otherwise breastfeeding in the current open interval. Alternatively one could take the most recently weaned child for every woman.

[75] For a systematic overview of the problems of relating cumulated fertility experience to any intermediate fertility variables or to socio-economic characteristics that are not constant over time, when these variables are measured only for the moment of the survey or for a very recent period, see Leridon 1980.

7 Summary and Conclusions

The definitions used can have a major impact on analysis of WFS breastfeeding data. Two major points should be reiterated here. First, exclusion of currently pregnant women (as was implicit, for example, in the WFS definition of each woman's current open birth interval) can lead to significant under-representation of women with short birth intervals, who tend, in many populations, to be women with shorter than average breastfeeding. In Pakistan, exclusion of pregnant women would lead us to overestimate the average duration of breastfeeding by about six months. Pregnant women should, therefore, not be excluded from analyses. Secondly, the focus on the most recent ('last') closed birth interval or on the current open birth interval for each woman, which was implicit in the restriction of detailed questions to at most the two most recent births per woman, can also give rise to results that are extremely hard to interpret. Exclusive use of data from the last closed birth interval can result in quite severe under-representation of long birth intervals, which often translates into a marked under-representation of women with long breastfeeding: this is especially so for women who are at short marriage durations. Again, in Pakistan the resulting bias could be as large as about six months. Data from the current open birth interval can also be biased, the bias this time being in favour of longer than average birth intervals, although its magnitude is usually smaller. Furthermore, the various selection biases are not equally severe for all subgroups in the population, and the presence of different biases may jeopardize attempts to analyse differences in actual behaviour. Finally, we should reiterate the point that data for the two most recent births per woman do not refer to the same period of time for all women. Where breastfeeding patterns have been changing, this lack of a common time reference can undermine any analysis based on these data.

It is clear that the most straightforward way to estimate breastfeeding patterns and differentials from WFS data is to analyse the breastfeeding information for all children born in a given period preceding the survey, rather than analysing the data for the two most recent births per woman. The period concerned should be defined so as to be as long as the longest duration of breastfeeding in the population. In those few countries where direct questions on breastfeeding were asked for *all* births in a sufficiently long period before the survey, full life-table methods can be employed to develop a breastfeeding table. This is possible, for example, in countries like Pakistan that asked questions about breastfeeding for all births, and also for countries that only asked about two births but where breastfeeding is so short that no woman could have had more than two births during the period covered. For all other countries we can use information on current breastfeeding status either to estimate a breastfeeding table or simply to estimate just the mean duration of breastfeeding.

Examination of the impact of breastfeeding on fertility raises additional problems over and above those already mentioned. Since for most countries the measure of breastfeeding used as the independent variable must usually come from data for the last closed birth interval, there is no entirely clean method of analysis. The selection biases in the breastfeeding data are not necessarily the same as the biases in the dependent variable selected. This is most patently evident if we attempt to relate length of breastfeeding in the last closed birth interval to the length of that interval itself. The average length of that interval is usually more biased than the average length of any of its components. The potential problems are usually less if we relate breastfeeding to data on amenorrhoea in the same interval, or if we take each woman's breastfeeding in that interval as a proxy for her average breastfeeding to date and relate that to her total fertility experience to date (parity). Even here, however, the problems are not entirely absent. We have, at the very least, to spell out exactly what data we have used and avoid unnuanced interpretation of the results obtained. Often we underestimate breastfeeding's impact.

A second issue that arises in analysis of the impact of breastfeeding on fertility is the choice of fertility measure used. Unfortunately, the core questionnaire contained no information on the two variables that capture the mechanisms through which breastfeeding exerts its impact on fertility: post-partum amenorrhoea and post-partum abstinence. In most countries we can estimate the number of months post-partum non-susceptibility to conception that is due to breastfeeding only by assuming that the crude average relationship found between breastfeeding and amenorrhoea elsewhere is valid also for our population and its various subgroups. If we want to analyse the relationship between breastfeeding and actual fertility *levels* rather than fertility *processes*, we must use a measure such as birth interval length or parity. For any of these, the impact of breastfeeding may be either reinforced or compensated by the simultaneous impact of other variables. Multivariate analysis is then needed to attempt to disentangle the role of each.

Finally we should note that most WFS questionnaires did not include enough questions to permit the development and testing of a full behavioural model of the determinants of breastfeeding practices. In the present analysis we have restricted ourselves to merely an analysis of the differentials related to a very few background socio-economic variables (age, education, residence, etc) that are available everywhere. The general forms of multivariate analysis we have used could, however, be applied for quantifying and testing a behavioural model wherever suitable data exist.

On the substantive side we have seen that breastfeeding in Pakistan is both nearly universal and quite prolonged, with an estimated 98.5 per cent of all surviving children (94.3 per cent of all children) receiving breastmilk and the average duration of breastfeeding being estimated as 21.8 months for surviving children (19.2 months for all children). Children born to younger women (or to urban residents) are breastfed less than the others. Those born to women with post-primary education are breastfed quite markedly less, but the mother's exposure to education up to but not beyond the primary level appears to have little effect on breastfeeding.

Breastfeeding plays a significant role in fertility in Pakistan. If the relationship between the average durations of breastfeeding and amenorrhoea found in certain other populations also prevails in Pakistan, then breastfeeding adds about 11 months on average to the period of non-susceptibility to conception after each birth. Breastfeeding is also a major correlate of differences in parity, its influence being offset only to a slight extent by the influence of other variables, most notably contraception.

Overall the picture is one of breastfeeding practices that are still relatively intact in Pakistan as a whole. Breastfeeding traditions are eroded, however, among the younger and the most modernized segments of the population, leading to slightly more rapid childbearing in these groups. At the time of the Pakistan Fertility Survey (1975), the proportions who breastfeed less were still quite small. They are likely to increase, however, as modernization and urbanization proceed.

References

Akin, J., R. Bilsborrow, D. Guilkey, B.M. Popkin, D. Benoit, P. Cantrelle, M. Garenne and P. Levi (1981). The Determinants of Breast-Feeding in Sri Lanka. *Demography 18(3)*: 287-307.

Barlow, R.E., D.J. Bartholomew, J.M. Bremner and H.D. Brunk (1972). *Statistical Inference under Order Restrictions.* New York: Wiley.

Beyer, W.H. (1966). *Handbook of Tables for Probability and Statistics.* Ohio: Chemical Rubber Co.

Bongaarts, J. (1981). The Proximate Determinants of Natural Marital Fertility. Draft of paper for US National Academy of Science, *Determinants of Fertility in Developing Countries: a Summary of Knowledge* (forthcoming).

Bonnar, J., N. Franklin, P.N. Nott and A.S. McNeilly (1975). Effect of Breast-Feeding on Pituitary-Ovarian Function after Childbirth. *British Medical Journal 1975(1)*: 82-4.

Bonte, M. and H. Van Balen (1969). Prolonged Lactation and Family Spacing in Rwanda. *Journal of Biosocial Science 1*: 97-100.

Boulier, B. and M. Rosenzweig, M. (1978). Age, Biological Factors and Socio-Economic Determinants of Fertility: a New Measure of Cumulative Fertility for Use in the Empirical Analysis of Family Size. *Demography 15(4)*: 487-97.

Braun, H.I. (1980). Regression-like Analysis of Birth Interval Sequences. *Demography 17*: 207-23.

Caldwell, J.C. and P. Caldwell (1977). The Role of Marital Sexual Abstinence in Determining Fertility: a Study of the Yoruba of Nigeria. *Population Studies 31(2)*: 193-217.

Caldwell, P. and J.C. Caldwell (1981). The Function of Child-Spacing in Traditional Societies and the Direction of Change. In H.J. Page and R. Lesthaeghe, eds *Child-Spacing in Tropical Africa: Traditions and Change.* London: Academic Press.

Cantrelle, P. and B. Ferry (1979). Approche de la fécondité naturelle dans les populations contemporaines. In H. Leridon and J. Menken, eds *Natural Fertility – Fécondité Naturelle.* International Union for the Scientific Study of Population.

Caraël, M. (1981). Child-Spacing, Ecology and Nutrition in the Kivu Province of Zaire. In H.J. Page and R. Lesthaeghe, eds *Child-Spacing in Tropical Africa: Traditions and Change.* London: Academic Press.

Chen, L.C., S. Ahmed, M. Gesche and W.H. Mosley (1974). A Prospective Study of Birth Interval Dynamics in Rural Bangladesh. *Population Studies 28(2)*: 277-96.

Chidambaram, V.C., J.G. Cleland and V. Verma (1980). Some Aspects of WFS Data Quality: a Preliminary Assessment. *WFS Comparative Studies* no 16.

Chidambaram, V.C. and T.W. Pullum (1981). Estimating Fertility Trends from Retrospective Birth Histories: Sensitivity to Imputation of Missing Dates. *Population Studies 35(2)*: 307-20.

Coale, A.J., A.G. Hill and T.J. Trussell (1975). A New Method of Estimating Standard Fertility Measures from Incomplete Data. *Population Index 41(2)*: 182-210.

Cole, E. (1979). Breast-Feeding: a Critique of the Literature. In Dana Raphael, ed *Breast-Feeding and Food Policy in a Hungry World.* London: Academic Press.

Corsini, C. (1979). Is the Fertility-Reducing Effect of Lactation Really Substantial? In H. Leridon and J. Menken, eds *Natural Fertility – Fécondité Naturelle.* International Union for the Scientific Study of Population.

Cox, D.R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society, Series B, 34*: 187-202.

Delvoye, P., J. Delogne-Desnoeck and C. Robyn (1976). Serum Prolactin in Long-Lasting Lactational Amenorrhoea. *Lancet ii*: 288.

Delvoye, P., M. Demaegd, J. Delogne-Desnoeck and C. Robyn (1977). The Influence of the Frequency of Nursing and of Previous Lactation Experience on Serum Prolactin in Lactating Mothers. *Journal of Biosocial Science 9*: 447-52.

Ferry, B. (1981). Breastfeeding. *WFS Comparative Studies* no 13.

Henry, L. (1972). *On the Measurement of Human Fertility: Selected Writings.* Translated by M.C. Sheps and E. Lapierre-Adamcyk. Elsevier.

Howie, P.W. (1981). Effect of Supplementary Food on Suckling Patterns and Ovarian Activity during Lactation. *British Medical Journal 283*: 757.

Howie, P.W. and A.S. McNeilly (1982a). Effect of Breastfeeding on Human Birth Intervals. *Journal of Reproduction and Fertility* (in press).

Howie, P.W. and A.S. McNeilly (1982b). Breastfeeding and Post-Partum Ovulation. *IPPF Medical Bulletin 16(2)*: 1–3.

Huffman, S.L., A.K.M.A. Chowdhury, J. Chakraburty and W.H. Mosley (1978). Nutrition and Post-partum Amenorrhoea in Rural Bangladesh. *Population Studies 32(2)*: 251-60.

Jain, A.K. and J. Bongaarts (1981). Breast-Feeding: Patterns, Correlates and Fertility Effects. *Studies in Family Planning 12(3)*: 79-99.

Jain, A.K., A. Hermalin and T.H. Sun (1979). Lactation and Natural Fertility. In H. Leridon and J. Menken, eds

Natural Fertility – Fécondité Naturelle. International Union for the Scientific Study of Population.

Kay, R. (1977). Proportional Hazard Regression Models and the Analysis of Censored Survival Data. *Applied Statistics 26*: 227-37.

Knodel, J. and N. Debavalya (1980). Breast-Feeding in Thailand: Trends and Differentials, 1969-79. *Studies in Family Planning 11(12)*: 355-77.

Konner, M. and C. Worthman (1980). Nursing Frequency, Gonadal Functions and Birth Spacing among Kung Hunter-Gatherers. *Science 207*: 788-91.

Leridon, H. (1977). *Human Fertility: the Basic Components.* Chicago University Press.

Leridon, H. (1980). L'interprétation des histoires génésiques quand les données sur les variables intermédiaires ou explicatives sont limitées à la situation au moment de l'enquête. Paper prepared for the seminar, The Analysis of Maternity Histories, International Union for the Scientific Study of Population/World Fertility Survey/London School of Hygiene and Tropical Medicine (forthcoming).

Lesthaeghe, R., P. Ohadike, J. Kocher and H.J. Page (1981). Child-Spacing and Fertility in Sub-Saharan Africa: an Overview of Issues. In H.J. Page and R. Lesthaeghe, eds *Child-Spacing in Tropical Africa: Traditions and Change.* London: Academic Press.

Lesthaeghe, R. and H.J. Page (1980). The Post-Partum Non-Susceptible Period: Development and Application of Model Schedules. *Population Studies 34(1)*: 143-69.

Little, R.J.A. (1978). Generalized Linear Models for Cross-Classified Data from the WFS. *WFS Technical Bulletins* no. 5.

McDonald, J. (1981). A New Methodological Approach for the Analysis of WFS Current Status Breast-Feeding Data. WFS Technical Paper no 1732.

McNeilly, A.S. (1979). Effects of Lactation on Fertility. *British Medical Bulletin 35*: 151.

Masnick, G.S. (1979). The Demographic Impact of Breast-Feeding: a Critical Review. *Human Biology 51(2)*: 109-25.

Menken, J., T.J. Trussell, D. Stempel and O. Babakol (1981). Proportional Hazards Life Table Models: an Illustrative Analysis of Socio-Demographic Influences on Marriage Dissolution in the United States. *Demography 18(2)*: 181-200.

Mosley, W.H., L.H. Werner and S. Becker (1982). The Dynamics of Birth Spacing and Marital Fertility in Kenya. *WFS Scientific Reports* no 30.

Namboodiri, N.K., L.F. Carter and H.M. Blalock (1975). *Applied Multivariate Analysis and Experimental Designs.* Maidenhead: McGraw-Hill.

Page, H.J., B. Ferry, I.H. Shah and R. Lesthaeghe (1980). The Most Recent Births; Analytical Potential and Some Underlying Problems. Paper prepared for the seminar, The Analysis of Maternity Histories, International Union for the Scientific Study of Population/World Fertility Survey/London School of Hygiene and Tropical Medicine (forthcoming).

Page, H.J. and R. Lesthaeghe (eds) (1981). *Child-Spacing in Tropical Africa: Traditions and Change.* London: Academic Press.

Pakistan Population Planning Council (1976). *Pakistan Fertility Survey: First Report.*

Perez, A., P. Vela, R.G. Potter and G.S. Masnick (1971). Timing and Sequence of Resuming Ovulation and Menstruation after Childbirth. *Population Studies 25*: 491-503.

Popkin, B.M., R. Bilsborrow, M. Yamamoto and J. Akin (1979). Breast-Feeding in Low Income Countries: Patterns and Determinants. *Carolina Population Center Papers* no 11. Chapel Hill.

Potter, R.G., M. New, J. Wyon and J. Gordon (1965). Applications of Field Studies to Research on the Physiology of Human Reproduction. In M.C. Sheps and J.C. Ridley, eds *Public Health and Population Change.* University of Chicago Press.

Rodríguez, G. and J.N. Hobcraft (1980). Illustrative Analysis: Life Table Analysis of Birth Intervals in Colombia. *WFS Scientific Reports* no 16.

Rolland, R., F.H. De Jong, L.A. Schellekens and R.M. Leguin (1975). The Role of Prolactin in the Restoration of Ovarian Function during Early Post-Partum Period in the Human Female: a Study during Inhibition of Lactation by Bromergocryptine. *Clinical Endocrinology 4*: 27-38.

Salber, E.J., M. Feinleib and B. MacMahon (1966). The Duration of Postpartum Amenorrhoea. *American Journal of Epidemiology 82(3)*: 347-58.

Santow, G. (1978). *A Simulation Approach to the Study of Human Fertility.* Martinus Nijhof.

Schoenmaeckers, R., I.H. Shah, R. Lesthaeghe and O. Tambashe (1981). Child-Spacing in Tropical Africa: Anthropological Evidence. In H.J. Page and R. Lesthaeghe, eds *Child-Spacing in Tropical Africa: Traditions and Change.* London: Academic Press.

Sheps, M.C. and J.A. Menken (1973). *Mathematical Models of Conception and Birth.* Chicago University Press.

Singarimbum, M. and C. Manning (1976). Breast-Feeding, Amenorrhoea and Abstinence in a Javanese Village: a Case Study in Mojolama. *Studies in Family Planning 7(7)*: 175-9.

Smith, D.P. (1980). Life Table Analysis. *WFS Technical Bulletins* no 6.

Smith, D.P. (1981). Regression Analysis of 'Current Status' Life Tables on Duration of Breast-Feeding. WFS Technical Paper no 1690.

Tukey, J.W. (1977). *Exploratory Data Analysis.* London: Addison-Wesley.

Tyson, J.E., R.S. Freedman, A. Perez, H.A. Zacur and J. Zanurtu (1976). Significance of the Secretion of Human Prolactin and Gonadotropin for Puerperal Lactational Infertility. *Breast-Feeding and the Mother.* Ciba Foundation Symposium 45: 49-71. Excepta Medica, Elsevier.

Tyson, J.E. and A. Perez (1978). The Maintenance of Infecundity in Postpartum Women. In W.H. Mosley, ed *Nutrition and Human Reproduction.* London: Plenum.

Ware, H. (1977). Language Problems in Demographic Field Work in Africa: the Case of the Cameroon Fertility Survey. *WFS Scientific Reports* no 2.

Appendix A — Pakistan Fertility Survey: Breastfeeding Rates

NOTE: The mean duration cited in the tables is the mean for those who were breastfed.
All tables in appendix A were computed using the FORTRAN program LACTATE developed by David Smith.

Table A1 Proportions ever breastfed and proportions currently being breastfed, by age: children born in the five years preceding interview

Women aged 15–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	254.	230.	.96602	.90520	.01839
6	299.	254.	.93441	.84966	.02068
9	286.	239.	.95251	.83683	.02187
12	357.	272.		.76376	.02251
15	190.	129.		.67784	.03398
18	253.	141.	MEAN =	.55390	.03126
21	225.	96.	.95012	.42572	.03299
24	362.	99.		.27282	.02343
27	190.	41.		.21429	.02979
30	283.	37.		.12866	.01991
33	244.	25.		.10225	.01940
36	401.	9.		.02133	.00722
39	243.	3.		.00989	.00636
42	257.	3.		.00997	.00620
45	243.	3.		.00989	.00636
48	402.	0.		0.00000	0.00000
51	214.	0.		0.00000	0.00000
54	247.	0.		0.00000	0.00000
57	241.	0.		0.00000	0.00000
60	419.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 19.76 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 19.26 MONTHS

Women aged 15–24 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	119.	104.	.94356	.87085	.03077
6	122.	100.	.89727	.82257	.03469
9	129.	110.	.95906	.84657	.03175
12	154.	112.		.72783	.03598
15	61.	38.		.61339	.06238
18	94.	44.	MEAN =	.46462	.05161
21	86.	28.	.93373	.32736	.05089
24	152.	34.		.22085	.03369
27	71.	11.		.15618	.04338
30	92.	7.		.06730	.02621
33	76.	7.		.08874	.03281
36	129.	2.		.00929	.00845
39	90.	0.		0.00000	0.00000
42	89.	0.		0.00000	0.00000
45	68.	0.		0.00000	0.00000
48	111.	0.		0.00000	0.00000
51	65.	0.		0.00000	0.00000
54	83.	0.		0.00000	0.00000
57	70.	0.		0.00000	0.00000
60	119.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 18.26 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 17.29 MONTHS

Women aged 25–34 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	101.	96.	.98101	.94408	.02290
6	138.	120.	.96229	.87115	.02856
9	111.	88.	.93103	.79194	.03857
12	162.	129.		.79543	.03172
15	97.	68.		.70705	.04642
18	115.	69.	MEAN =	.59924	.04578
21	114.	54.	.95777	.47321	.04682
24	165.	48.		.29092	.03542
27	94.	21.		.21629	.04265
30	138.	11.		.07957	.02309
33	116.	10.		.07980	.02521
36	193.	4.		.01953	.00998
39	114.	2.		.01053	.00957
42	119.	2.		.01587	.01149
45	118.	2.		.01018	.00926
48	201.	0.		0.00000	0.00000
51	111.	0.		0.00000	0.00000
54	120.	0.		0.00000	0.00000
57	118.	0.		0.00000	0.00000
60	210.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 20.00 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 20.36 MONTHS

Women aged 35–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	35.	32.	1.00000	.91003	.04895
6	40.	35.	.95264	.85793	.05545
9	46.	43.	.98515	.91801	.04055
12	42.	32.		.77288	.06518
15	33.	23.		.71271	.07987
18	45.	29.	MEAN =	.62387	.07228
21	26.	15.	.97862	.53984	.09784
24	46.	18.		.38124	.07221
27	27.	10.		.36111	.09341
30	55.	20.		.35679	.06511
33	54.	10.		.16999	.05145
36	80.	4.		.04524	.02333
39	39.	2.		.03088	.02778
42	50.	1.		.01364	.01642
45	58.	2.		.02093	.01893
48	91.	0.		0.00000	0.00000
51	39.	0.		0.00000	0.00000
54	44.	0.		0.00000	0.00000
57	53.	0.		0.00000	0.00000
60	91.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 22.09 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 21.75 MONTHS

Table A2 Proportions ever breastfed and proportions currently being breastfed, by age: children born in the five years preceding interview – surviving children

Women aged 15–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	236.	230.	.99134	.97471	.01023
6	267.	254.	.98264	.95086	.01323
9	259.	239.	.99013	.92227	.01664
12	306.	272.		.88957	.01793
15	163.	129.		.79064	.03195
18	215.	141.	MEAN =	.65310	.03250
21	193.	96.	.98789	.49669	.03603
24	297.	99.		.33257	.02737
27	160.	41.		.25548	.03458
30	232.	37.		.15756	.02397
33	201.	25.		.12468	.02336
36	309.	9.		.02769	.00934
39	204.	3.		.01178	.00757
42	212.	3.		.01208	.00751
45	197.	3.		.01221	.00784
48	302.	0.		0.00000	0.00000
51	176.	0.		0.00000	0.00000
54	195.	0.		0.00000	0.00000
57	193.	0.		0.00000	0.00000
60	348.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 21.58 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 20.94 MONTHS

Women aged 15–24 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	109.	104.	.98741	.95746	.01942
6	106.	100.	.96289	.94511	.02217
9	119.	110.	.99427	.91954	.02497
12	128.	112.		.87181	.02958
15	52.	38.		.72278	.06225
18	75.	44.	MEAN =	.58022	.05707
21	73.	28.	.98207	.38272	.05700
24	123.	34.		.27241	.04015
27	56.	11.		.19706	.05340
30	74.	7.		.08355	.03226
33	54.	7.		.12437	.04508
36	93.	2.		.01291	.01172
39	73.	0.		0.00000	0.00000
42	73.	0.		0.00000	0.00000
45	52.	0.		0.00000	0.00000
48	75.	0.		0.00000	0.00000
51	53.	0.		0.00000	0.00000
54	66.	0.		0.00000	0.00000
57	56.	0.		0.00000	0.00000
60	93.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 20.04 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 19.22 MONTHS

Women aged 25–34 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	97.	96.	.99295	.98590	.01201
6	126.	120.	1.00000	.95267	.01893
9	98.	88.	.98067	.90333	.02999
12	145.	129.		.88923	.02610
15	86.	68.		.79936	.04344
18	103.	69.	MEAN =	.67128	.04645
21	98.	54.	.99194	.55098	.05033
24	137.	48.		.35112	.04090
27	84.	21.		.24200	.04693
30	109.	11.		.10103	.02897
33	100.	10.		.09298	.02916
36	155.	4.		.02433	.01240
39	97.	2.		.01241	.01127
42	99.	2.		.01902	.01375
45	96.	2.		.01250	.01135
48	155.	0.		0.00000	0.00000
51	91.	0.		0.00000	0.00000
54	95.	0.		0.00000	0.00000
57	93.	0.		0.00000	0.00000
60	177.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 21.49 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 21.77 MONTHS

Women aged 35–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	32.	32.	1.00000	1.00000	0.00000
6	36.	35.	.98077	.96154	.03234
9	44.	43.	1.00000	.97231	.02496
12	34.	32.		.95916	.03430
15	26.	23.		.89947	.05962
18	38.	29.	MEAN =	.74910	.07089
21	23.	15.	.99380	.63114	.10241
24	38.	18.		.46345	.08174
27	21.	10.		.47051	.11080
30	50.	20.		.39271	.06963
33	48.	10.		.19149	.05721
36	62.	4.		.05844	.02992
39	35.	2.		.03470	.03116
42	41.	1.		.01666	.02004
45	50.	2.		.02435	.02198
48	73.	0.		0.00000	0.00000
51	33.	0.		0.00000	0.00000
54	35.	0.		0.00000	0.00000
57	45.	0.		0.00000	0.00000
60	80.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 25.12 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 23.35 MONTHS

Table A3 Proportions ever breastfed and proportions currently being breastfed, by age: children born in the five years preceding interview – surviving children of rural residents

Women aged 15–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	168.	167.	1.00000	.99286	.00651
6	214.	210.	.99438	.98315	.00882
9	186.	174.	.99355	.93548	.01804
12	238.	221.		.92929	.01665
15	113.	93.		.81915	.03629
18	163.	115.	MEAN =	.70588	.03571
21	139.	77.	.99577	.55172	.04220
24	211.	81.		.38068	.03345
27	111.	32.		.28261	.04291
30	172.	33.		.18881	.02991
33	144.	22.		.15000	.02979
36	222.	8.		.03243	.01190
39	147.	3.		.01639	.01051
42	154.	2.		.00781	.00711
45	133.	3.		.01802	.01154
48	215.	0.		0.00000	0.00000
51	121.	0.		0.00000	0.00000
54	141.	0.		0.00000	0.00000
57	142.	0.		0.00000	0.00000
60	253.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 22.57 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 21.91 MONTHS

Women aged 15–24 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	76.	75.	1.00000	.98413	.01439
6	84.	82.	.98571	.97143	.01820
9	88.	83.	1.00000	.94521	.02435
12	101.	91.		.90476	.02927
15	35.	26.		.72414	.07586
18	56.	36.	MEAN =	.65217	.06419
21	51.	21.	.99515	.40476	.06923
24	91.	29.		.31579	.04873
27	42.	10.		.22857	.06488
30	54.	5.		.08889	.03878
33	36.	6.		.16667	.06219
36	69.	2.		.01754	.01589
39	54.	0.		0.00000	0.00000
42	54.	0.		0.00000	0.00000
45	34.	0.		0.00000	0.00000
48	54.	0.		0.00000	0.00000
51	36.	0.		0.00000	0.00000
54	51.	0.		0.00000	0.00000
57	42.	0.		0.00000	0.00000
60	74.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 20.81 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 19.85 MONTHS

Women aged 25–34 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	66.	66.	1.00000	1.00000	0.00000
6	100.	99.	1.00000	.98795	.01095
9	62.	56.	.98039	.90196	.03806
12	112.	105.		.93548	.02328
15	60.	51.		.84000	.04739
18	78.	56.	MEAN =	.70769	.05156
21	72.	45.	.99471	.61667	.05737
24	94.	39.		.41026	.05091
27	58.	17.		.29167	.05996
30	80.	10.		.12121	.03672
33	71.	8.		.10169	.03597
36	111.	3.		.02174	.01390
39	66.	2.		.01818	.01647
42	70.	2.		.01724	.01562
45	66.	2.		.01818	.01647
48	107.	0.		0.00000	0.00000
51	63.	0.		0.00000	0.00000
54	64.	0.		0.00000	0.00000
57	65.	0.		0.00000	0.00000
60	126.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 22.58 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 22.70 MONTHS

Women aged 35–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	27.	27.	1.00000	1.00000	0.00000
6	30.	30.	1.00000	1.00000	0.00000
9	38.	36.	1.00000	.96774	.02900
12	26.	26.		1.00000	0.00000
15	18.	17.		.93333	.05887
18	30.	24.	MEAN =	.80000	.07312
21	17.	12.	1.00000	.71429	.11035
24	27.	14.		.50000	.09743
27	11.	5.		.44444	.15139
30	39.	18.		.46875	.08063
33	38.	9.		.22581	.06864
36	44.	4.		.08333	.04210
39	27.	2.		.04545	.04059
42	30.	0.		0.00000	0.00000
45	34.	2.		.03571	.03206
48	54.	0.		0.00000	0.00000
51	23.	0.		0.00000	0.00000
54	27.	0.		0.00000	0.00000
57	35.	0.		0.00000	0.00000
60	54.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 26.16 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 24.00 MONTHS

Table A4 Proportions ever breastfed and proportions currently being breastfed, by age: children born in the five years preceding interview – surviving children of urban residents

Women aged 15–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	68.	64.	.97000	.93000	.03094
6	54.	45.	.93421	.82278	.05210
9	74.	66.	.98148	.88889	.03667
12	69.	52.		.75248	.05208
15	50.	37.		.72603	.06330
18	52.	26.	MEAN =	.48684	.06953
21	54.	20.	.96479	.35443	.06526
24	86.	19.		.21429	.04433
27	49.	10.		.19444	.05656
30	60.	5.		.06818	.03258
33	57.	4.		.06024	.03167
36	88.	2.		.01563	.01329
39	58.	0.		0.00000	0.00000
42	59.	2.		.02326	.01971
45	64.	0.		0.00000	0.00000
48	88.	0.		0.00000	0.00000
51	56.	0.		0.00000	0.00000
54	55.	0.		0.00000	0.00000
57	51.	0.		0.00000	0.00000
60	95.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 18.72 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 17.83 MONTHS

Women aged 15–24 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	33.	30.	.95833	.89583	.05347
6	22.	19.	.87500	.84375	.07784
9	32.	27.	.97826	.84783	.06422
12	28.	21.		.75000	.08303
15	17.	13.		.72000	.10890
18	20.	8.	MEAN =	.37931	.10927
21	23.	8.	.94444	.33333	.09951
24	32.	5.		.14894	.06298
27	14.	2.		.10000	.08135
30	20.	2.		.06897	.05706
33	18.	1.		.03846	.04574
36	25.	0.		0.00000	0.00000
39	19.	0.		0.00000	0.00000
42	19.	0.		0.00000	0.00000
45	18.	0.		0.00000	0.00000
48	22.	0.		0.00000	0.00000
51	17.	0.		0.00000	0.00000
54	15.	0.		0.00000	0.00000
57	14.	0.		0.00000	0.00000
60	20.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 17.78 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 16.94 MONTHS

Women aged 25–34 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	31.	30.	.97778	.95556	.03725
6	27.	22.	1.00000	.82051	.07452
9	37.	33.	.98113	.90566	.04869
12	34.	25.		.73469	.07648
15	26.	18.		.70270	.09112
18	25.	14.	MEAN =	.55556	.10043
21	26.	10.	.98507	.36842	.09489
24	43.	10.		.22222	.06352
27	26.	4.		.13158	.06650
30	30.	2.		.04651	.03894
33	29.	3.		.07143	.04819
36	45.	2.		.03077	.02598
39	31.	0.		0.00000	0.00000
42	30.	1.		.02326	.02787
45	30.	0.		0.00000	0.00000
48	49.	0.		0.00000	0.00000
51	28.	0.		0.00000	0.00000
54	32.	0.		0.00000	0.00000
57	28.	0.		0.00000	0.00000
60	51.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 18.46 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 18.89 MONTHS

Women aged 35–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	5.	5.	1.00000	1.00000	0.00000
6	6.	5.	.87500	.75000	.18565
9	7.	7.	1.00000	1.00000	0.00000
12	9.	7.		.83333	.13046
15	8.	7.		.81818	.14102
18	8.	5.		.54545	.18206
21	6.	3.		.37500	.20757
24	11.	5.		.37500	.14677
27	10.	5.		.50000	.16205
30	11.	2.		.12500	.10026
33	11.	1.		.06667	.07810
36	19.	0.		0.00000	0.00000
39	9.	0.		0.00000	0.00000
42	11.	1.		.06250	.07339
45	16.	0.		0.00000	0.00000
48	19.	0.		0.00000	0.00000
51	11.	0.		0.00000	0.00000
54	9.	0.		0.00000	0.00000
57	10.	0.		0.00000	0.00000
60	26.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 21.24 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 18.80 MONTHS

Table A5 Proportions ever breastfed and proportions currently being breastfed, by age: children born in the five years preceding interview – surviving children, both parents illiterate

Women aged 15–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	133.	132.	.99488	.98975	.00874
6	153.	149.	.98762	.97101	.01357
9	143.	134.	.99524	.93426	.02074
12	158.	150.		.94890	.01757
15	86.	72.		.83947	.03980
18	113.	71.	MEAN =	.62734	.04563
21	103.	63.	.99242	.60318	.04825
24	164.	62.		.37373	.03782
27	86.	26.		.30067	.04967
30	134.	26.		.18984	.03397
33	124.	19.		.14612	.03172
36	171.	6.		.03507	.01408
39	100.	2.		.01198	.01088
42	117.	3.		.02138	.01354
45	114.	2.		.01057	.00961
48	178.	0.		0.00000	0.00000
51	102.	0.		0.00000	0.00000
54	112.	0.		0.00000	0.00000
57	108.	0.		0.00000	0.00000
60	218.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 22.67 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 22.35 MONTHS

Women aged 15–24 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	59.	58.	.98836	.97671	.01973
6	46.	45.	.97396	.97396	.02349
9	60.	57.	.98865	.93464	.03193
12	62.	59.		.94976	.02793
15	30.	23.		.78257	.07656
18	37.	24.	MEAN =	.64351	.07939
21	38.	16.	.98444	.42289	.08060
24	62.	16.		.25693	.05557
27	30.	8.		.26297	.08052
30	38.	1.		.01818	.02184
33	34.	7.		.19919	.06905
36	47.	0.		0.00000	0.00000
39	32.	0.		0.00000	0.00000
42	37.	0.		0.00000	0.00000
45	30.	0.		0.00000	0.00000
48	38.	0.		0.00000	0.00000
51	26.	0.		0.00000	0.00000
54	43.	0.		0.00000	0.00000
57	28.	0.		0.00000	0.00000
60	55.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 21.07 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 19.95 MONTHS

Women aged 25–34 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	50.	50.	1.00000	1.00000	0.00000
6	83.	81.	1.00000	.96908	.01903
9	51.	46.	1.00000	.91463	.03951
12	76.	72.		.94359	.02652
15	43.	36.		.82803	.05774
18	54.	31.	MEAN =	.56978	.06738
21	53.	37.	1.00000	.68721	.06370
24	77.	33.		.42699	.05674
27	45.	12.		.25687	.06543
30	59.	10.		.16434	.04855
33	56.	6.		.09819	.03988
36	87.	3.		.02764	.01761
39	48.	0.		0.00000	0.00000
42	55.	2.		.03424	.02456
45	53.	0.		0.00000	0.00000
48	94.	0.		0.00000	0.00000
51	55.	0.		0.00000	0.00000
54	48.	0.		0.00000	0.00000
57	51.	0.		0.00000	0.00000
60	111.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 22.26 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 23.16 MONTHS

Women aged 35–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	25.	25.	1.00000	1.00000	0.00000
6	25.	24.	.97198	.97198	.03350
9	33.	32.	1.00000	.96351	.03273
12	21.	20.		.96627	.04021
15	14.	14.		1.00000	0.00000
18	22.	17.	MEAN =	.74255	.09349
21	13.	10.	.99169	.79205	.11574
24	26.	13.		.49684	.09846
27	11.	7.		.58734	.15019
30	38.	16.		.40015	.07990
33	35.	6.		.17175	.06389
36	38.	4.		.09600	.04817
39	21.	2.		.05744	.05097
42	26.	1.		.02637	.03155
45	32.	2.		.03790	.03398
48	48.	0.		0.00000	0.00000
51	23.	0.		0.00000	0.00000
54	22.	0.		0.00000	0.00000
57	30.	0.		0.00000	0.00000
60	53.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 26.64 MONTHS

Table A6 Proportions ever breastfed and proportions currently being breastfed, by age: children born in the five years preceding interview – surviving children, mother illiterate and father literate

Women aged 15–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	78.	75.	.98249	.95832	.02268
6	90.	84.	.98463	.94080	.02499
9	89.	83.	.98649	.93645	.02592
12	115.	100.		.87201	.03121
15	63.	49.		.76966	.05307
18	69.	50.	MEAN =	.72684	.05399
21	64.	24.	.98462	.36737	.06041
24	98.	34.		.34473	.04822
27	58.	11.		.18298	.05084
30	78.	10.		.11958	.03695
33	58.	7.		.10655	.04062
36	109.	2.		.01735	.01256
39	75.	2.		.01608	.01458
42	65.	0.		0.00000	0.00000
45	62.	2.		.01954	.01768
48	94.	0.		0.00000	0.00000
51	56.	0.		0.00000	0.00000
54	59.	0.		0.00000	0.00000
57	61.	0.		0.00000	0.00000
60	105.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 20.93 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 19.89 MONTHS

Women aged 15–24 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	33.	31.	.97915	.94245	.04078
6	45.	43.	.96950	.94265	.03482
9	46.	43.	1.00000	.93209	.03739
12	47.	40.		.85244	.05214
15	18.	13.		.68251	.11053
18	29.	16.	MEAN =	.54485	.09327
21	25.	9.	.98334	.33076	.09456
24	48.	17.		.34213	.06886
27	19.	2.		.10376	.07170
30	28.	5.		.15258	.06796
33	16.	0.		0.00000	0.00000
36	36.	2.		.03390	.03046
39	33.	0.		0.00000	0.00000
42	20.	0.		0.00000	0.00000
45	16.	0.		0.00000	0.00000
48	28.	0.		0.00000	0.00000
51	21.	0.		0.00000	0.00000
54	17.	0.		0.00000	0.00000
57	22.	0.		0.00000	0.00000
60	29.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 19.38 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 18.63 MONTHS

Women aged 25–34 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	40.	39.	.98275	.96550	.02907
6	35.	33.	1.00000	.94114	.03998
9	33.	31.	.96363	.92232	.04666
12	56.	48.		.87026	.04525
15	34.	27.		.79308	.06969
18	31.	28.	MEAN =	.89347	.05597
21	31.	11.	.98234	.36184	.08741
24	42.	14.		.32525	.07273
27	31.	6.		.18109	.07008
30	40.	2.		.03450	.02907
33	33.	4.		.11519	.05592
36	54.	1.		.01277	.01539
39	32.	2.		.03810	.03415
42	33.	0.		0.00000	0.00000
45	30.	2.		.04056	.03631
48	47.	0.		0.00000	0.00000
51	26.	0.		0.00000	0.00000
54	34.	0.		0.00000	0.00000
57	28.	0.		0.00000	0.00000
60	52.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 21.34 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 20.22 MONTHS

Women aged 35–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
3	6.	6.	1.00000	1.00000	0.00000
6	10.	10.	1.00000	.93133	.08037
9	11.	11.	1.00000	1.00000	0.00000
12	14.	13.		.94825	.06112
15	12.	10.		.83571	.10962
18	10.	7.	MEAN =	.74040	.14437
21	9.	5.	1.00000	.49064	.16944
24	9.	4.		.45848	.17413
27	10.	4.		.33902	.15320
30	10.	4.		.36979	.15491
33	10.	3.		.24582	.13797
36	20.	0.		0.00000	0.00000
39	11.	0.		0.00000	0.00000
42	13.	0.		0.00000	0.00000
45	17.	0.		0.00000	0.00000
48	19.	0.		0.00000	0.00000
51	11.	0.		0.00000	0.00000
54	9.	0.		0.00000	0.00000
57	12.	0.		0.00000	0.00000
60	25.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 23.58 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 20.89 MONTHS

Table A7 Proportions ever breastfed and proportions currently being breastfed, by age: children born in the five years preceding interview – surviving children, mother with primary education

Women aged 15–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
6	31.	29.	1.00000	.95467	.03798
12	40.	31.		.79039	.06512
18	40.	26.		.64988	.07600
24	32.	9.	MEAN =	.28274	.08041
30	28.	6.	1.00000	.18395	.07348
36	33.	1.		.02101	.02521
42	33.	0.		0.00000	0.00000
48	37.	0.		0.00000	0.00000
54	24.	0.		0.00000	0.00000
60	27.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 20.30 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 20.45 MONTHS

Women aged 15–24 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
6	20.	20.	1.00000	1.00000	0.00000
12	24.	19.		.81025	.08112
18	11.	6.		.55055	.15554
24	14.	3.		.19223	.10804
30	12.	3.		.21570	.12346
36	10.	0.		0.00000	0.00000
42	14.	0.		0.00000	0.00000
48	10.	0.		0.00000	0.00000
54	4.	0.		0.00000	0.00000
60	6.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 19.61 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 18.85 MONTHS

Women aged 25–34 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
6	9.	8.	1.00000	.84026	.12556
12	16.	12.		.76083	.10767
18	23.	16.		.65716	.09930
24	14.	6.		.41319	.13339
30	14.	2.		.10100	.08212
36	19.	1.		.03597	.04283
42	15.	0.		0.00000	0.00000
48	22.	0.		0.00000	0.00000
54	15.	0.		0.00000	0.00000
60	19.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 19.85 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 21.87 MONTHS

Women aged 35–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
6	2.	2.	1.00000	1.00000	0.00000
12	0.	0.		0.00000	0.00000
18	7.	5.		.78450	.16367
24	5.	1.		.15336	.17112
30	4.	2.		.42014	.27434
36	5.	0.		0.00000	0.00000
42	5.	0.		0.00000	0.00000
48	7.	0.		0.00000	0.00000
54	6.	0.		0.00000	0.00000
60	3.	0.		0.00000	0.00000

Table A8 Proportions ever breastfed and proportions currently being breastfed, by age: children born in the five years preceding interview – surviving children, mother with post-primary education

Women aged 15–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
6	15.	11.	.86364	.72727	.11515
12	18.	10.		.54347	.11901
18	19.	5.		.26158	.10303
24	23.	2.		.08488	.05927
30	14.	0.		0.00000	0.00000
36	21.	0.		0.00000	0.00000
42	22.	0.		0.00000	0.00000
48	19.	0.		0.00000	0.00000
54	15.	0.		0.00000	0.00000
60	21.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 14.24 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 12.93 MONTHS

Women aged 15–24 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	PROPORTIONS: CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
6	9.	5.	.83333	.58333	.17259
12	7.	3.		.40982	.19090
18	9.	2.		.15674	.12342
24	8.	2.		.16359	.13675
30	6.	0.		0.00000	0.00000
36	8.	0.		0.00000	0.00000
42	7.	0.		0.00000	0.00000
48	7.	0.		0.00000	0.00000
54	5.	0.		0.00000	0.00000
60	5.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 12.46 MONTHS

Women aged 25–34 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
6	7.	7.	.90000	.90000	.11504
12	11.	7.		.62500	.14677
18	9.	4.		.38462	.16363
24	15.	1.		.04596	.05443
30	9.	0.		0.00000	0.00000
36	11.	0.		0.00000	0.00000
42	15.	0.		0.00000	0.00000
48	7.	0.		0.00000	0.00000
54	10.	0.		0.00000	0.00000
60	13.	0.		0.00000	0.00000

MEAN DURATION OF BREAST-FEEDING = 16.04 MONTHS
MEDIAN DURATION OF BREAST-FEEDING = 15.12 MONTHS

Women aged 35–49 at event

AT X= (MONTH)	SAMPLE SIZE	NUMBER BREAST-FEEDING	PROPORTIONS: EVER BREASTFED [=L(0)]	CURRENT [=L(X)]	S.E. OF L(X) UNDER S.R.S.
6	0.	0.	0.00000	0.00000	0.00000
12	0.	0.		0.00000	0.00000
18	1.	0.		0.00000	0.00000
24	0.	0.		0.00000	0.00000
30	0.	0.		0.00000	0.00000
36	3.	0.		0.00000	0.00000
42	0.	0.		0.00000	0.00000
48	6.	0.		0.00000	0.00000
54	2.	0.		0.00000	0.00000
60	4.	0.		0.00000	0.00000

Appendix B – Standard Breastfeeding Schedule

Table B1 Standard schedule of breastfeeding

Duration (in months) (d)	Standard expressed as logit $P_s(d)$ $Y_s(d)$	Standard expressed as proportions still breastfeeding at d $P_s(d)$	Standard expressed as proportions weaning between d – 1 and d $w_s(d-1, d)$
0	∞	1.000	
1	1.92	0.979	0.021
2	1.72	0.969	0.010
3	1.54	0.956	0.013
4	1.37	0.939	0.017
5	1.20	0.917	0.022
6	1.03	0.887	0.030
7	0.87	0.851	0.036
8	0.71	0.805	0.046
9	0.55	0.750	0.055
10	0.39	0.686	0.064
11	0.23	0.614	0.072
12	0.08	0.540	0.074
13	–0.07	0.465	0.075
14	–0.22	0.392	0.073
15	–0.36	0.327	0.065
16	–0.50	0.269	0.058
17	–0.64	0.218	0.051
18	–0.78	0.174	0.044
19	–0.92	0.137	0.037
20	–1.06	0.107	0.030
21	–1.20	0.083	0.024
22	–1.33	0.065	0.018
23	–1.46	0.051	0.014
24	–1.59	0.040	0.011
25	–1.72	0.031	0.009
26	–1.85	0.024	0.007
27	–1.97	0.019	0.005
28	–2.09	0.015	0.004
29	–2.21	0.012	0.003
30	–2.33	0.009	0.003
31	–2.45	0.007	0.002
32	–2.56	0.006	0.001
33	–2.67	0.005	0.001
34	–2.78	0.004	0.001
35	–2.89	0.003	0.001
36	–2.99	0.003	<0.001
37	–3.09	0.002	–
38	–3.19	0.002	–
39	–3.29	0.002	–
40	–3.38	0.001	–
41	–3.48	0.001	–
42	–3.57	<0.001	–
43	–3.65	–	–
44	–3.73	–	–
45	–3.81	–	–

Source: Lesthaeghe and Page 1980.