PN-AAL-309

ISN: 13869

AMERICAN PUBLIC HEALTH ASSOCIATION
International Health Programs
1015 Fifteenth Street, NW
Washington, DC 20005

AN ANALYSIS OF EXTANT SURVEY AND CENSUS DATA: PROBLEMS AND SOLUTIONS

A Report Prepared By:
JAMES F. PHILLIPS, PH.D.

During the Period:
MAY 2-13, 1982

Supported By The:
U.S. AGENCY FOR INTERNATIONAL DEVELOPMENT
(ADSS) AID/DSPE-C-0053

AUTHORIZATION:
Ltr. AID/DS/POP: 8-9-82
Assgn. No. 582144

ACKNOWLEDGMENTS

The author gratefully acknowledges the technical assistance he received throughout this consultancy from Dr. M. Nizamuddin, United Nations Fund for Population Activities Adviser to the Department of Statistics, Hashemite Kingdom of Jordan, and the logistics and clerical support provided by the USAID mission in Amman.

C O N T E N T S

Page

ACKNOWLEDGMENTS i
ABBREVIATIONS v

I. INTRODUCTION AND BACKGROUND 1

Scope of Work 1
Itinerary 2
The Setting 3
The Problem 3

II. OBSERVATIONS AND FINDINGS 5

The Components of Fertility 5
Possible Biases to Explain Apparent Discrepancy Between Fertility Levels and Contraceptive Use 10

A. Evidence of Fertility Overreporting: Fertility Time Reference Errors 10
B. Family Planning Response Errors 10

III. RECOMMENDATIONS 19

Questionnaires for Proposed Studies 19
Analysis of Extant Data 20

IV. CONCLUSIONS 23

BIBLIOGRAPHY 25

APPENDICES

Appendix A: List of Persons Contacted

Appendix B: Technical Assistance Report to USAID/Amman on Obstacles to Data Analysis at Selected Health and Population Science Research Centers

Appendix C: Report on Technical Assistance to the Department of Community Medicine, University of Jordan, on the Health Status in the Jordan Survey of 1980

PREVIOUS PAGE BLANK

ABBREVIATIONS

AIT	Asian Institute of Technology
APHA	American Public Health Association
BSHSJ	Baseline Study of the Health Status of Jordan
CBR	Crude Birth Rate
CDR	Crude Death Rate
CE	Contraceptive Effectiveness
CPS	Contraceptive Prevalence Survey
DOS	Department of Statistics
GOJ	Government of Jordan
ICDDR	International Centre for Diarrhoeal Disease Research (Bangladesh)
JFS	Jordan Fertility Survey
MWRA	Married Women of Reproductive Age
NAS	National Acceptor Survey
NFS	National Fertility Survey
NPC	National Planning Council
OBI	Open Birth Interval
PHC	Primary Health Care
POPLAB	International Program of Laboratories for Population Statistics (North Carolina)
TFR	Total Fertility Rate
UDD	Urban Development Department
WFS	World Fertility Survey
WHS	Westinghouse Health Systems
UNFPA	United Nations Fund for Population Activities
USAID	United States Agency for International Development

I. INTRODUCTION AND BACKGROUND

I. INTRODUCTION AND BACKGROUND

Scope of Work

The scope of work for this consultancy called for an analysis of extant survey and census data. Some surveys were cited in a brief statement of work in progress, and, in the original scope of work, technical assistance was sought in the broad area of finalizing reports. The scope of work was changed considerably on the first day of the consultancy.[1]

Because this was to be a brief consultancy, the United States Agency for International Development (USAID) requested a general assessment of why analysis of certain studies seems to be proceeding slowly, despite the rather substantial commitment of resources for data collection and analysis. The primary issue to be addressed was not an analysis of data, but an analysis of why research was progressing so slowly in a number of major studies, and what could be done to remedy the situation. One study, the Baseline Study of the Health Status of Jordon (BSHSJ), was singled out for special attention because its problems are not well understood and AID's involvement to date has been substantial.

Unpublished data could not be analyzed in the time available because computer software is limited, facilities have not been fully developed, and data sets are incomplete in some cases. Although it would have been useful to tabulate data from the Jordan Fertility Survey (JFS), counterparts from the Department of Statistics (DOS) were engaged in an external review that left them little time to collaborate in the consultancy for the American Public Health Association (APHA).

These conditions detracted from the goal of analyzing Jordanian demographic data, but it was possible, nevertheless, to address certain aspects of the planned scope of work during the consultancy; namely, the question of why fertility is so high when prevalence of contraceptive use is also high, and what the implications of this question are for future demographic research. This question, considered by both DOS and USAID staff, became the central issue addressed during the assignment.

Work consisted of three activities: discussion with scientists and staff at various research institutions and USAID officials about the barriers to analysis of extant data; the provision of technical assistance to the staff of the Department of Community Medicine, where data management problems are perceived to be especially complex and where investigators have expressed an interest in exchanging ideas; and discussions with DOS and AID staff on JFS fertility data and the implications of the findings for a next round of a demographic survey.

---

[1] A list of the persons contacted during the consultancy appears in Appendix A.

The analysis of extant data, one aspect of the scope of work that was formally agreed upon by the DOS in advance of the consultancy, was the theme of the assignment and is thus the focus of this report. The first two activities described above were treated as separate issues, conducted in response to USAID requests, and have, therefore, been consigned to Appendices B and C. Issues were discussed as numbered topics. Topic 1, for example, concerned the evaluation of computer facilities; the study of this subject, which was somewhat peripheral to the APHA consultant's area of expertise, will require a carefully defined joint scope of work for consultants and Jordanian computer managers. Nevertheless, in response to the USAID's request, the APHA consultant made certain observations and concluded that expert assessment of the computer science needs of Jordanian research institutions merits support, if a request for such a review emanates from the National Planning Council (NPC). A review of the consultant's informal discussions and recommendations appears in Appendix B.

The second activity was the result of exchanges with the staff of the Department of Community Medicine. After a day of informal discussions, the investigators requested a more formal presentation, including written notes and a group discussion of the recommendations. A synopsis of that interchange appears in Appendix C.

Itinerary

The consultancy was located in Amman from May 2, 1982, to May 13, 1982. The first work day was devoted to an orientation at the USAID mission and the Department of Statistics, and to a review of documents and reports. On the second day, the consultant met the staff of the Department of Community Medicine and attended orientation sessions on the Baseline Survey of the Health Status of the Jordanian Population. The third day was spent reviewing documents from the survey and formulating preliminary impressions and recommendations. This activity was followed by a review with the investigators of some written comments and suggestions (see Appendix C).

Following initial discussions and reviews, the consultant began to analyze published data from the Jordan Fertility Survey and to discuss his impressions and conclusions with Dr. M. Nizamuddin of the United Nations Fund for Population Activities (UNFPA), a technical adviser to the Department of Statistics. With the help of the DOS, further discussions of the data and future research plans were arranged with the faculty of the University of Jordan's Department of Demography. General research problems and plans were discussed with DOS staff during the two final days of the consultancy. Logistics support and clerical assistance was provided by the USAID mission. All work was conducted in close consultation with Mr. Scott Edmunds, a program officer at the mission, and Dr. Nizamuddin.

The Setting

The Hashemite Kingdom of Jordan, British Transjordan, a political entity dating to 1923, was established in 1950. The population of the occupied East Bank was approximately 900,000 at the time of occupation--1967; it is thought that several hundred thousand persons fled to the West Bank.

The East Bank population has grown exceedingly rapidly since 1967, from approximately 500,000 at the time of independence to nearly two million by 1975, when work on the Jordan Fertility Survey began. One estimate places the population of the Kingdom at 3,189,000 for mid-1979. Growth rates in the Kingdom are high, 3.8 - 3.9 annually, because of high fertility (the crude birth rate [CBR] is 49/1,000) and low mortality (the crude death rate [CDR] is 11/1,000).[2] The crude death rate has fallen considerably in recent years because of improvements in public health. The population is young, with 50 percent under the age of 15. The prevalence of contraceptive use is also high.

Jordan is an arid country. Accordingly, geographic distribution is concentrated in areas where there is access to water. Approximately one-half of the population resides in Amman, the capital.

Literacy in Jordan is high among males--82 percent; female literacy is somewhat lower--59 percent. Currently, enrollment is nearly universal; thus, this differential will dissipate in time.

Jordan does not have a national population program, although a modest private clinical family planning program and commercial sales of contraception are allowed. Knowledge of contraception is nearly universal.

The Problem

Fertility in Jordan is high. Various estimates since 1970 place the total fertility rate (TFR) above 7. The 1972 National Fertility Survey (NFS) estimate is 7.60;[3] the 1976 Jordan Fertility Survey estimate is 7.15;[4] and the estimate of the Urban Development Department (UDD), which conducted the Baseline Health and Population Assessment in 1980, is 7.40.[5]

---

[2] See estimates of the U.S. Bureau of the Census, Department of Commerce, 1980.

[3] See Osmun Nur, 1981, page 139.

[4] See Department of Statistics, 1979.

[5] See the report of the Population Council and the Urban Development Department.

These estimates are consistent with estimates of fertility levels based on the approximations of stable population or calculations using the indirect estimation method of Coale, Hill, and Trussell.[6] There is little doubt that fertility in Jordan has been high over the past 10 years, and probably for the two preceding decades as well.

The fertility rate estimated in the 1976 JFS is not surprising, but it appears to be inconsistent with the survey's reported levels of contraceptive use in the Kingdom. Of the women who are exposed to the risk of pregnancy (women who are currently married, fecund, and not pregnant), some 37 percent were contracepting at the time of the survey. Thus, this question arises: How can fertility and contraceptive use be concomitantly high? Are these findings mutually inconsistent? If so, what indirect evidence is there that fertility is overreported, that contraceptive use is overrepresented, or both? And, in any case, what are the research implications of the findings?

In addressing these questions, the consultant attempted, first, to ascertain whether the rates are, in fact, inconsistent, given the JFS data on the proximate components of fertility and the efficacy of contraception, and then to examine possible biases in the data.

---

[6] See Coale et al., 1975.

II. OBSERVATIONS AND FINDINGS

II. OBSERVATIONS AND FINDINGS

The Components of Fertility

In their pioneering work on fertility determinants, Davis and Blake (1956) showed that all variation in human fertility can be explained by variation in 11 intermediate variables, of which contraceptive practice is but one. To give a perspective to the question of how concomitantly high fertility and contraceptive practice can arise, the components of fertility were analyzed.

In a recent exegesis of the Davis and Blake framework, Bongaarts (1975) noted that the fecundability rate is approximately 15.4 and that natural fertility is, therefore, potentially extremely high. Although prevalence of contraceptive use is important from a policy standpoint, contraceptive practice may be a relatively minor component of the proximate determinants of fertility in a given population. Bongaarts (1978) identified four components which merit attention in decompositions of fertility determinants: proportions married, which he defines as the proportion of women of reproductive age (MWRA) who engage in sexual activity regularly (no distinction is made between formal marriages and consensual unions); abortion, which is defined as any practice that deliberately interrupts gestation; lactational infecundability, which is defined as the period during which women are unable to conceive in the postpartum period until the normal pattern of ovulation and menstruation resumes; and contraceptive effectiveness, the extent to which fertility is reduced by contraceptive practice. Bongaarts (1978) proposed the following relationship between these components and the total fertility rate:

$$TFR = C_m \, C_a \, C_i \, C_c \, TF \qquad \text{..... (1)}$$

where TF is the total fecundability rate and subscripts m, a, i, and c index the proportionate reduction in fertility from the prevailing pattern and level of marriage, abortion, infecundability, and contraceptive use, respectively. Osman Nur estimated each component in his recent dissertation on the quality of the 1976 JFS data. The following quotation is taken from Chapter 5 of Dr. Nur's dissertation:[7]

---

[7] Osmun Nur, "On the Validity of Fertility Analysis for Jordan," Unpublished Ph.D. Dissertation, University of Michigan, 1981.

"Except for contraceptive effectiveness (e), most of the data required to estimate the fertility-inhibiting effects for Jordan (i.e., the "C" values) can be obtained from the 1972 NFS and the 1976 JFS. Since induced abortion is not practiced in Jordan, the index of abortion "C_a" is equated to unity. The fertility-reducing impact of contraception "C_c" is estimated by the following equation:

$$C_c = 1 - sue \qquad (5.2)$$

"where

"u = average proportion of married women currently using contraception (average of age-specific use rates);

"e = average contraceptive effectiveness (average of use-effectiveness levels by age and method);

"s = sterility correction factor (according to Bongaarts, s = 1.1 for developing countries).

"In most of the developing countries, the data needed to estimate the average contraceptive effectiveness is lacking. Using the 1974 Philippine National Acceptor Survey, Laing was able to estimate the rates of contraceptive effectiveness for each method. Since effectiveness depends on the method, we obtained an overall measure of contraceptive effectiveness "e" for Jordan by weighting the method-specific contraceptive effectiveness for the Philippines, the weights being the proportions currently using each method given in the 1972 NFS and the 1976 JFS.

"Table 1 summarizes reproductive measures and indices of the behavioral intermediate fertility variables obtained from the reproduction information in the 1972 NFS and the 1976 JFS.[8] Several observations can be made: 1) During 1972 and 1976, the indices of proportion married and not contracepting [sic] have declined. However, the decline for the latter index is significantly lower than the former, suggesting that nuptiality change is a more important factor in the recent fertility decline compared to contraception use. 2) The index of lactational infecundability increased slightly from 0.73 to 0.75 as a result of a decline in lactation during 1972 and 1976. Lactation has an inhibitory effect on ovulation and thus increases the birth interval and reduces "natural fertility." The small change in the

---

[8] Table 1 in Dr. Nur's dissertation is Table 5.5. The Laing reference is in Studies in Family Planning, Vol. 9, No. 6, 1978.

"Table 1: Recorded and Estimated Reproductive Measures and Indices of Intermediate Fertility Variables (Jordan 1972 and 1976)

Measures and Indices	1972	1976	1976/1972
Proportion Currently Using Contraception "u"	0.21	0.24	1.14
Contraceptive Effectiveness "e"[a]	0.81	0.80	0.99
Lactational Infecundability (i)[b]	0.90	0.83	0.92
Index of Proportion Married "C_m"	0.76	0.70	0.92
Index of Non-Contraception "C_c"	0.81	0.79	0.98
Index of Lactational Infecundability "C_i"	0.73	0.75	1.03
Combined Index $C_m \times C_c \times C_i$	0.449	0.415	0.92
Estimated TFR Using Bongaarts' Model	6.91	6.39	0.92
Recorded TFR	7.60	7.15	0.94

a Contraceptive effectiveness (e) is obtained by weighting the method-specific contraceptive effectiveness for the Philippines (in 1974), the weights being the proportions currently using each method (obtained from the 1972 NFS 1976 JFS).

b Average duration of lactational infecundability is estimated from the mean duration (in months) of lactation, "L", with the following equation: $i = 1.5 + 0.56L$.

"index of lactational infecundability indicates that the fertility-prompting effect of the decline in lactation during 1972 and 1976 is not significant. Nevertheless, the decline in the index is consistent with the recent trend in the decline of breastfeeding in developing countries. 3) The estimates of TFR by Bongaarts' model are less than the observed rates in 1972 NFS and 1976 JFS. However, the disagreement is not substantial and could partly be attributed to the overestimation of the fertility-reducing effect of contraception as a result of exaggerating current use of contraception commonly noticed in the 1972 NFS and the 1976 JFS. Since effectiveness of contraception (e) depends on the method, the value for the Philippines may be overestimating the measure for Jordan. In addition, errors in the specification of the model and errors in the observed total fertility rates may lead to disagreement between the observed TFR and the estimated TFR."

Thus, it is possible to conclude from Dr. Nur's application of the Bongaarts' model that high levels of contraceptive use (24 percent prevalence) and concomitantly high fertility are to be expected. The fact that the observed fertility level is 10.6 higher than levels predicted in Dr. Nur's computations may be attributable perhaps to Dr. Nur's assumptions about the inefficiency of contraceptive methods (i.e., use-effectiveness). It should be noted that Laing, in an article published after that cited by Nur, concluded that the effectiveness of the rhythm method in the Philippines is higher than was anticipated because of continuation rates among the Filipinos who use that method.[9] Nur's C_c might be spuriously inflated by his application of Philippine values to the Jordanian situation.

In a recent study of Jordanian urban women,[10] Bongaarts' decomposition was applied to users of efficient methods only. In the sample population, the TFR was 7.4, and the projection was 7.1, based on the assumption that 14 percent of women are contracepting (see Table II-1). Thus, the level of efficient method use implied a level of fertility roughly equivalent to the levels observed in the sample. The results suggest that inefficient contraception is, indeed, efficient in Jordan. Prevalence is, therefore, best represented by citing prevalence of efficient methods only.

---

9 See Laing, 1980.

10 See A Baseline Health and Population Assessment for the Upgrading Areas of Amman: A Report to the Municipality of Amman, Urban Development Department, Amman, March 1982.

Table II-1

FERTILITY-INHIBITING EFFECTS OF VARIOUS TYPES OF BEHAVIOR

Fertility Components	Children per Woman
Standard Total Fecundity Rate	15.3
Reductions Due To:	
1. Breastfeeding averages 11.5 months, causing an average postpartum period of amenorrhea, estimated at 5.8 months.	2.75
2. Contraception is currently practiced by 14 percent of married women with moderately effective methods.	1.76
3. The fertility-weighted proportion of women not exposed to childbearing because they are in a non-married state is 34 percent.	3.67
Total Fertility Rate after Deductions	7.1
Recorded TFR	7.4

Source: Population Council and the Urban Development Department, 1982.

Possible Biases to Explain Apparent Discrepancy Between Fertility Levels and Contraceptive Use

Fertility is approximately 10 percent higher than expected, given other proximate determinants. This discrepancy disappears when use is attributed to use of efficient methods only. Because it is unlikely that the impact of inefficient contraception is inconsequential, an unexplained discrepancy remains. This discrepancy could arise from biases in the data, such as overreporting of fertility, overreporting of contraceptive use, or both. Or it could arise from overestimations of the efficiency of use. For this assignment, only published JFS data were available. In examining the data, a search was made for evidence of Types 1, 2, or 3 biases, as shown in Table II-2. A Type-1 situation would contribute most to the observed discrepancy.

A. Evidence of Fertility Overreporting: Fertility Time Reference Errors

Recent period fertility may be slightly overreported from surveys because of errors in reference periods. Respondents tend to push events forward in time that occurred in the distant past. This tendency is illustrated in Figure II-1 by the period age-specific marital fertility rates for four periods. The pattern for recent period fertility is plausible, but among women 35 and older who are asked to recall events in the period 5-9 years before the survey, the rates are implausibly high, probably because older women report events as having occurred closer to the survey than was actually the case. Osman Nur (1981) analyzed this subject in detail and concluded that misplacement of events renders the JFS useless for the analysis of trends, but that recent period fertility is not seriously biased. Nevertheless, there is a possibility that misplacement of events has inflated recent current estimates of fertility.

Nur analyzed the consistency of recent period fertility in the JFS and concluded that there is no evidence that the survey underreported data; moreover, Nur found considerable consistency across studies. From this, one could conclude that there is little evidence that biases in fertility measurement account for the observed discrepancy.

B. Family Planning Response Errors

If recent ever-users who have discontinued respond as current users, contraceptive use will be spuriously inflated. This problem was discussed with staff of the Department of Statistics, and minor modifications of the proposed JFS2 questionnaire were recommended so that probes can accompany current questions on contraceptive use. How these probes should be conducted is a matter to be pursued in pretesting.

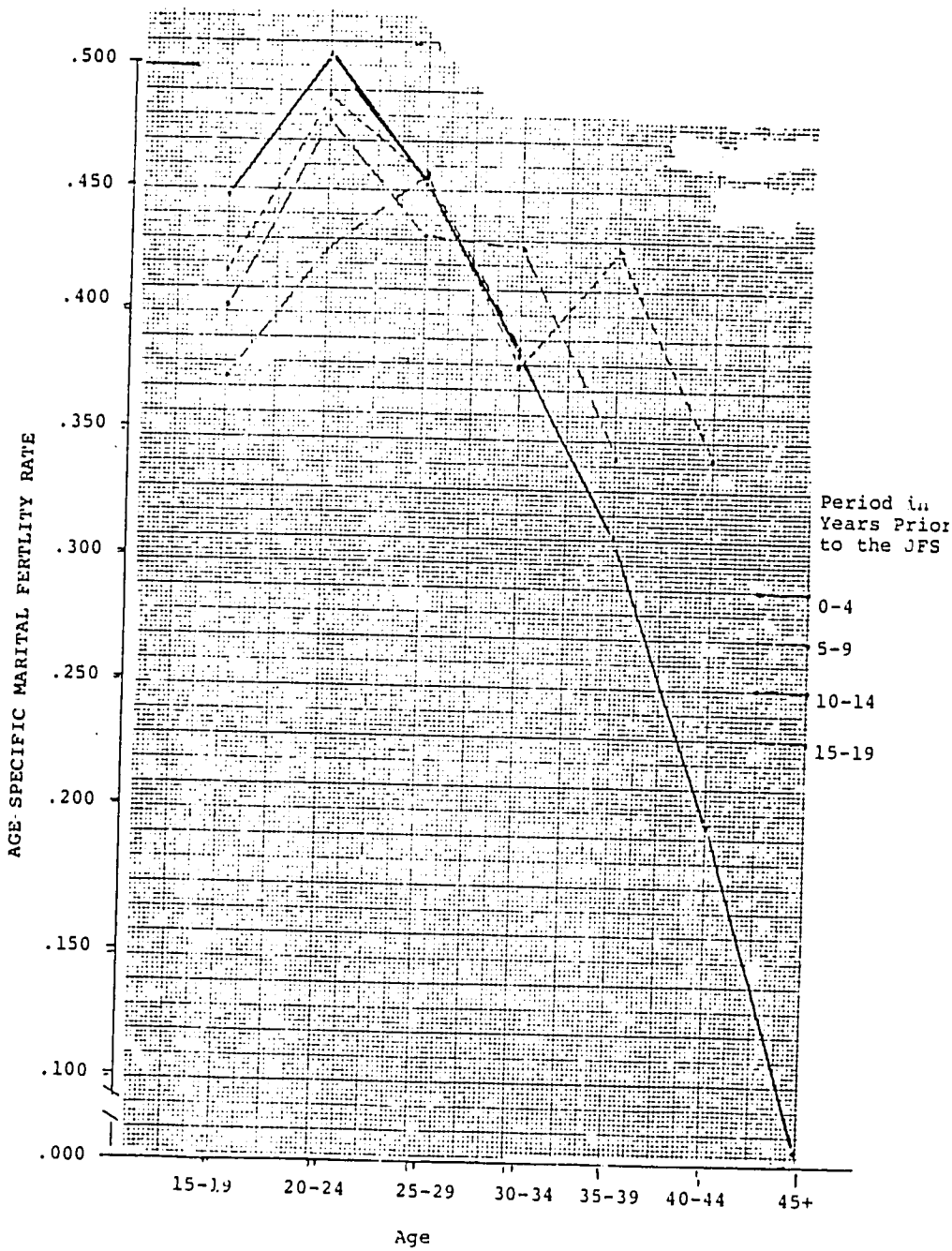
Table II-2

A TYPOLOGY OF BIASES IN LEVELS OF FERTILITY AND CONTRACEPTIVE USE PREVALENCE

	Biases that Lead to Fertility.....	
Biases Leading to Contraceptive Prevalence.....	Overreporting	Underreporting
Overreporting	(1)	(2)
Underreporting	(3)	(4)

Figure II-1

PERIOD MARITAL FERTILITY RATES
FOR FOUR FIVE-YEAR PERIODS,
1975 JFS



Interestingly, some useful probes existed in the 1976 JFS, but they do not appear in published tabulations, perhaps because responses were not meaningful.[11] The current use question is:

Q505: Are you or your husband currently using a method to keep you from getting pregnant?

Two untabulated probes are:

Q419: Is any pill available in this house now?

Q420: Are any condoms available in this house now?

Several useful tabulations are given in Table II-3. A similar table (see Table II-4) should be compiled for condoms.

Cells marked with an "x" would indicate responderts who may be recent users but who are not, in fact, current users. (Some respondents may be out of supplies and intend to replenish their contraceptives.) This table should be run to calculate the extent of the discrepancy.

Additional tabulations should be made to examine the relationship of ever-use and current use in the current open birth interval (OBI). The relationship between use of a particular method and open birth interval is illustrated in Fi e II-2. Current use begins approximately one year after partuiti although, in some cases, it begins earlier. Ever-use initially coincides with current use, but the two curves diverge with time because of discontinuation. If the gap between the two curves is narrow, use is efficient. If the gap between the curves is narrow for inefficient methods, such as withdrawal, one may suspect that ever-use is spuriously reported as current use. Nothing can be "proven" by drawing such graphs, but it is useful to compare the figures by method. (The requisite tabulations are not available in published form and could not be tabulated during the consultancy.)

It is appropriate to compare the Urban Development Department's questionnaire for the baseline health and population assessment with the JFS questionnaire. The UDD sample was entirely of the urban poor, and the results could not be expected to compare with the results of the JFS. What is impressive, however, is that fertility levels in both surveys compare closely, although contraceptive prevalence in the UDD's assessment is only 14 percent among the UDD respondents. Although inefficient use was not used in the decomposition in the UDD report, the decomposition of proximate fertility determinants is more plausible in

---

[11] JFS documents do not explain why the special probes have not been tabulated.

Table II-3

AVAILABILITY OF CONTRACEPTIVE PILL SUPPLIES,
BY METHOD CURRENTLY USED

Pill Supplies Available Now?	Method Used Now: Pill	Condom	....Others	None	Total
Yes					
No	X				
Total					

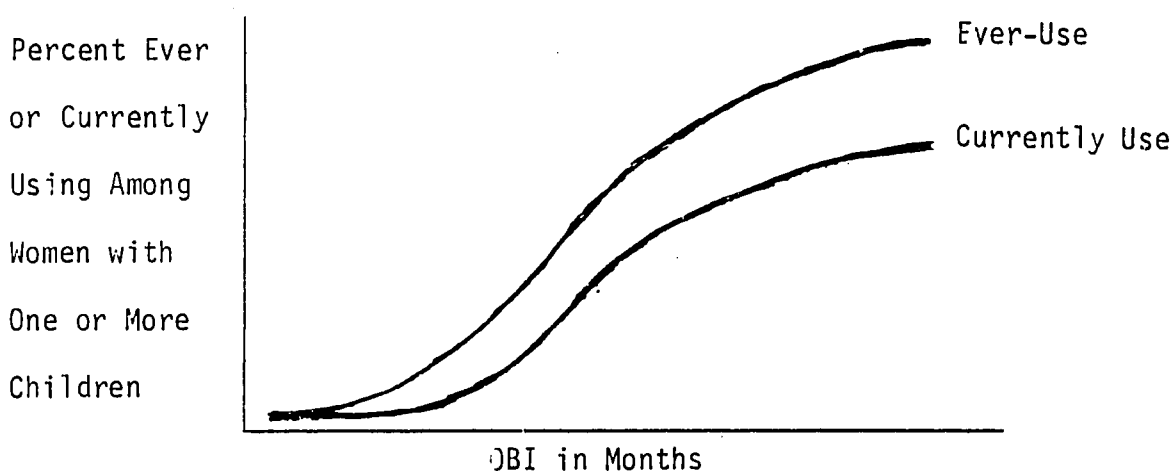
Table II-4

AVAILABILITY OF CONDOM SUPPLIES IN THE HOUSEHOLD NOW, BY METHOD CURRENTLY USED

Condom Supplies Available Now?	Pill	Condom	.....	Others	Total
Yes					
No		X			
Total					

Figure II-2

RELATIONSHIP BETWEEN PARTICULAR METHODS AND OBI



the UDD report than in the JFS, as was noted above. One might suggest that the UDD survey, which focused mainly on health and development issues, elicited more valid responses to questions on contraceptive use than did the JFS simply because the questions' salience in the former study was not obvious to the respondents and courtesy responses were thus not common.

Clearly, the methodological issues in contraceptive questioning in the Contraceptive Prevalence Survey (CPS) or the JFS merit careful investigation. In-depth evaluation of the concept of contraceptive prevalence questioning needs to be reviewed, perhaps by pretesting a small sample of contraceptive questions with an anthropological study of reactions to those questions and the validity of responses. The Department of Statistics might undertake such a study before the next JFS is fielded.

In summary, the examination of published extant data was inconclusive. The consultant suspects that the level of prevalence, or the efficiency of contraception, is overestimated in the JFS, particularly for the "inefficient methods." A plan of analysis was proposed to address this issue.

III. RECOMMENDATIONS

III. RECOMMENDATIONS

Current plans for the next Jordanian fertility survey call for one of two approaches: a contraceptive prevalence survey with simple indirect estimates of fertility, or a questionnaire designed by the World Fertility Survey (WFS) similar to the 1976 JFS. The recommendations included in this chapter cover the selection of a study that will produce results to questions about fertility determinants in Jordan and a proposed analysis of extant data.

Questionnaires for Proposed Studies

The Department of Statistics has proposed to undertake a fertility survey in 1983. Funds available from the USAID for survey research have been committed to Westinghouse Health Systems (WHS), an organization best known for contraceptive prevalence surveys. The DOS has requested a questionnaire similar to that used in the WFS, instead of the CPS, although Westinghouse may assist in a demographic survey. This APHA consultant endorses the DOS' position, in light of current questions about fertility determinants. The following factors support the position of the DOS:

1. The CPS questionnaire records extensive information about contraceptive prevalence, but little information that would permit assessment of the demographic impact of contraception. The questionnaire assesses "Brass fertility" measures, duration of breastfeeding, child expectations and desires, contraceptive knowledge and use, reasons for non-use, accessibility of services, method preferences, sources of supplies and preferences for supply sources, and a few background characteristics. Despite all the care and attention to contraception, the questionnaire does not address issues of the efficacy of use, because the proximate determinants of fertility are not assessed. Thus, the questions raised by the Government of Jordan (GOJ) and the USAID about the fertility impact of contraception and the relative importance of contraception among the proximate determinants cannot be answered in the CPS as it is structured at this time. The WFS instrument is better suited to answering policy questions, because it assesses complete fertility histories and proximate determinants.

2. Some information is needed, from a sample of women, on the use-effectiveness of contraception: duration of use in the open birth interval, termination of use as a result of method failure, and failure while using a method in the closed interval. Because questions are proposed in the WFS questionnaire that address the issue of effective use, C_c can be computed directly from Jordanian data, if that questionnaire is used. Some questions appear in the draft JFS2 questionnaire which may be helpful in solving this problem, but they require careful pretesting and validation.

3. The JFS has been shown to be inadequate for computing fertility trends, and there is evidence that recent fertility may be inflated slightly by displacement of events. Repeating an assessment of fertility history would permit a more complete evaluation of the quality of JFS data and would make possible an analysis of trends.

In summary, a WFS instrument is recommended, because the CPS approach does not permit assessment of the quality of fertility data or the effectiveness of contraceptives. The CPS is appropriate where policies are aimed at increasing contraceptive prevalence. In the absence of such a policy, it is appropriate to examine the broader questions of the determinants of fertility and the relative contribution of contraception to those determinants.

Analysis of Extant Data

Some tabulations of existing data are warranted. Although questions on fetal loss and abortion are likely to underreport true levels, the data collected in the JFS should be tabulated. The question of fetal loss was asked repeatedly, birth interval by birth interval. The responses should be tallied over all intervals and tabulated against parity. It should be noted that Dr. Osman Nur's assumption of "no abortion" was incorporated in Bongaarts' model calculations. This assumption could be examined by looking at the tabulations of JFS data on fetal loss.

"Contraceptive effectiveness" (CE) is an index of the proportionate decline of fertility while using a method. Actual calculation is somewhat complex; typically, it is based on some sample observations of users over time. Laing (Studies in Family Planning, 1978) computed CE from use-effectiveness data by examining the net conditional probability of pregnancy while each method is in use and relating that probability to the average population at risk to obtain an estimate of the fertility rate of users. This rate was then related to the expected fertility rate of the same users in the absence of contraception. Laing made these calculations by examining the potential fertility of fecund, married, non-contracepting women. The requisite data for this are:

1. Life table continuation rates for users;

2. Net failure rates;

3. The period of overlap between use and amenorrhea;

4. The proportion of failures ending in fetal loss; and

5. A fertility history for non-users who are fecund, married women matched to users by open birth interval.

Laing sampled from national service statistics records and interviewed retrospectively. His results provided invaluable insights into the efficacy of contraceptive use in the Philippines. In Jordan, analysts are forced to use Philippine CE data for every population where questions about the demographic impact of contraception arise. Because there is no frame in Jordan that is comparable to Laing's frame, some work with the next JFS is appropriate.

Some careful methodological work should be done to try to time contraceptive events in the last closed birth interval and the current open interval. One might conclude that reliable responses will not be forthcoming, but the issue needs to be explored. Currently, a question in the JFS2 about use in these intervals is asked. Information on how long use transpired, the timing of use in the postpartum period, and the last method used would provide the requisite data for Laing's analysis.

An alternative approach using extant data would be to analyze the prolongation of the birth interval associated with use in the last closed interval or current open interval. Such analysis can be done with JFS data. The approach is less elegant than Laing's approach, but estimation of a model is possible using limited JFS data. The basic model for the analysis would conform to the following specification:

$$\underset{\sim}{p} = \{1+e^{-[(\underset{\sim}{1}, \underset{\sim}{T}, \underset{\sim}{\chi})(\underset{\sim}{\alpha},\underset{\sim}{\beta},\underset{\sim}{\gamma})']}\}^{-1} \qquad .....(1)$$

$\underset{\sim}{p}$ = I x 1 vector of probabilities of a birth in time interval i, a parity progression probability;

$\underset{\sim}{1}$ = a vector of 1 in the data matrix;

$\underset{\sim}{T}$ = a J x 1 vector of time indicators conforming to some function defining the underlying risk of birth over time (Brass, 1981, proposes the Compertz model);

$\underset{\sim}{\chi}$ = a K x 1 vector of characteristics of women in the specified time intervals;

and,

$\underset{\sim}{\alpha}$ = a scalar;

$\underset{\sim}{\beta}$ = a set of unknown nuisance parameters for the components of time; and

$\underset{\sim}{\gamma}$ = a vector of unknown coefficients for the effect of covariates on the odds of having a birth. γ includes a coefficient for contraception. Its estimated value is a measure of CE.

Now (1) will be difficult to estimate in Jordan, given hardware, software, personnel, support services, and other constraints. Nevertheless, focusing upon the model in (1) would answer the basic question: To what extent does use of contraceptives affect fertility in Jordan? By expanding (1) to permit interactions between characteristics of women and use, one can examine the correlates of contraceptive effectiveness in Jordan.

The proposed analysis was discussed with Dr. Nizamuddin before the consultancy began, and it is considered to be a possible area of work once the requisite software has been installed at the DOS. Units of analysis are women-years of risk of birth. If feasible, a smaller time unit, such as three-month time segments, would be better. Analysis would be restricted to birth intervals in the JFS with contraceptive use data. Thus, $\underset{\sim}{X}$ would contain not only age and parity, but also dummy variables to indicate whether or not a given individual used contraception in the last closed interval or current open interval, with the dummies specifying whether or not use was "efficient" or "inefficient." In such an analysis γ_k would correspond to the effect of the kth covariate on the odds of a birth, such that the γ for contraception would represent indices of contraceptive effectiveness.

An analysis of this kind is planned for 1983 at the International Centre of Diarrhoeal Disease Research (ICDDR) in Bangladesh. The requisite software can be provided to the DOS on request. Although the model is seemingly complex, the programming requirements are quite simple, because standard, maximum-likelihood logistic-regression software will be incorporated in the statistical packages that are to be installed soon at the DOS.

A data matrix can be prepared by conducting rows for time units, with columns for the elements of $\underset{\sim}{T}$ and $\underset{\sim}{X}$. An example of this type of analysis appears in Phillips (1982),[12] although a more tractible approach is that of Menken et al. (1981). A useful adjunct to this work is a paper by Vanghan et al. (1980). Software for the Princeton analyses are not available in Amman at this time.

If this line of analysis is pursued, it should be conducted in collaboration with a research center with experience in the application of proportional hazards models to demographic data (e.g., Princeton's Office of Population Research or the International Program of Laboratories for Population Statistics (POPLAB) in North Carolina), because the techniques have yet to be demonstrated in Jordan.

---

[12] This analysis is "time-conditional" and thus more complex than the requisite analysis. The Menken approach is recommended.

IV. CONCLUSIONS

IV. CONCLUSIONS

The JFS offers considerable promise for addressing the question of the extent to which contraceptive use affects fertility in Jordan. The next JFS should build upon the first-round experience. Careful probing should be devised, so that use-prevalence responses in fact record current contraceptive practice. Even in the absence of precise data on the timing of use, however, much can be learned about contraceptive efficacy by proportional hazards analysis of birth interval dynamics, wherein covariates for use and non-use of contraception, by method, are included among the explanatory variables.

In Chapter I, several questions that arose from the scope of work for this consultancy were raised. Each of these questions is considered below.

1. How can fertility and contraceptive use be concomitantly high?

Proximate determinants other than contraception favor high fertility. The proportion married is sufficient to produce a C_m index that is comparable to levels observed among more traditional societies elsewhere in the developing world. Marital dissolution, which is common in some Islamic countries, does not appreciably reduce Jordanian fertility. Only approximately 7 percent of the women in the JFS sample had experienced a dissolution of marriage, and half of them had remarried before the interview. Interestingly, significant changes in nuptiality occurred between 1972 and 1976, and the observed trends tended to reduce fertility. The trends should be carefully monitored, because there is considerable potential for fertility change through changes in patterns of marriage. The contribution of nuptiality change to fertility trends over the period 1972 to 1976 was greater than the contribution of increased contraceptive usage.

Fecundability in Jordan is high; thus, C_i is high. Lactational amenorrhea is short, intrauterine mortality is low, and reproductive efficiency is high prior to adoption of a contraceptive method (see Nur, 1981). Thus, in certain respects Jordan is similar to developed countries: Natural fertility is high, so contraceptive prevalence must be very high for fertility to be low.

2. Are the findings on contraceptive use and fertility mutually inconsistent?

The data suggest that fertility is some 10 percent higher than one would predict from data on contraceptive use, amenorrhea,

and nuptiality. Perhaps a slightly higher discrepancy would be observed if abortion levels were known.

3. How does the discrepancy arise?

Recent period fertility may be slightly overreported because of the tendency of respondents to push events forward in time. Overreporting of these data primarily affects the estimation of fertility trends prior to the most recent period, and is not likely to explain the observed discrepancy. Nevertheless, it would be appropriate to reassess fertility histories in the next round to further analyze this problem.

More plausible is the hypothesis that contraceptive use and efficiency are overestimated. No research is available to examine this issue.

4. What are the research implications of these conclusions?

Some simple tabulations of extant JFS data are needed, in addition to proportional hazards analysis of birth interval dynamics. The latter analysis should be conducted in collaboration with an institution familiar with the techniques for such investigation.

In the future, data collection should proceed according to current DOS plans for a second round of the JFS, which would provide the requisite data for the analysis of fertility determinants.

BIBLIOGRAPHY

Bongaarts, John. "Why High Birth Rates are so Low." Population and Development Review, 1975, 1(2): 289-296.

_____. "A Framework for Analyzing the Proximate Determinants of Fertility." Population and Development Review, 1978, 4(1): 105-132.

Brass, W. "The Use of the Compertz Relational Model to Estimate Fertility." Paper Presented at the International Population Conference, International Union for the Scientific Study of Population, Manila, 1981.

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

Davis, Kingsley, and Judith Blake. "Social Structure and Fertility: An Analytical Framework." Economic Development and Social Change, 1956, 4: 211-235.

Department of Commerce. World Population 1979: Recent Demographic Estimates for the Countries and Regions of the World. Washington, D.C.: Bureau of the Census, United States Department of Commerce, 1980.

Department of Statistics, Hashemite Kingdom of Jordan. Jordan Fertility Survey, 1976. Vols. 1 and 2. London: Spotlswoode Ballantyne, Ltd., 1979.

Laing, J.E. "Estimating the Effects of Contraceptive Use on Fertility: Techniques and Findings from the 1974 Philippine National Acceptor Survey." Studies in Family Planning, 1978, 9(6): 150-162.

_____. "1976 National Acceptor Survey (NAS): Contraceptive Effectiveness of Methods Offered by the Philippine Family Planning Program." Research Note No. 143. Population Institute, University of Philippines System, 1980.

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

Nur, Osmur El-Hassan Mohammed. "On the Validity of Fertility Analysis for Jordan." Unpublished Ph.D. Dissertation. University of Michigan, 1981.

Phillips, J.F. "A Logit Regression Method for the Multivariate Analysis of Contraceptive Attrition." In The Role of Surveys in the Analysis of Family Planning Programs, ed. A.I. Hermaline and B. Entwistle, Liege, Belgium: Ordina Editions, 1982.

Population Council and the Urban Development Department. "A Baseline Health and Population Assessment for the Upgrading Areas of Amman: A Report to the Municipality of Amman." Amman, Urban Development Department, 1982.

Vaughan, B., J. Trussell, J. Menken, E.F. Jones, et al. "Contraceptive Efficacy among Married Women Aged 15-44 Years: United States, Hyattsville, MD." No. 5. DHHS Publication No. 80-1981. U.S. National Center for Health Statistics, 1980.

APPENDICES

Appendix A

LIST OF PERSONS CONTACTED

Appendix A

LIST OF PERSONS CONTACTED

United States Agency for International Development

Mr. Jack Thomas, Population Officer

Mr. Scott Edmunds, Program Officer

United Nations

Dr. Majeed Khan, Coordinator, UNFPA, Amman

Dr. Ranjan Kumar Som, Technical Adviser on Population, New York

Dr. M. Nizamuddin, Adviser, Department of Statistics, Amman

University of Jordan

Dr. Ahmed A. Hammouda, Associate Professor, Population Studies Department

Dr. Mohammed Galal Eldin, Associate Professor, Population Studies Department

Dr. Najwa Khuri-Bulos, Associate Professor, Pediatrics, University Medical Center

Dr. Sami A. Khoury, Dean, School of Nursing

Dr. Sa'ad S. Hijazi, Chairman, Department of Community Medicine

Dr. Adnan Abbas, Department of Community Medicine

Mrs. Salwa Masri

Dr. Amis Shant

Dr. Khalid Shami

Dr. Madi Wali

Dr. Suhail Barakat

Population Council

Dr. Laila Erader, Associate

Department of Statistics, Amman

Dr. Burhan Shrydeh, Director

Mr. Jawad Salah, Deputy Director

Mrs. Huda Ghneim Shbul, Head, Computer Department

Appendix B

TECHNICAL ASSISTANCE REPORT TO USAID/AMMAN ON OBSTACLES TO DATA ANALYSIS AT SELECTED HEALTH AND POPULATION SCIENCE RESEARCH CENTERS

Appendix B

TECHNICAL ASSISTANCE REPORT TO USAID/AMMAN ON OBSTACLES TO DATA ANALYSIS AT SELECTED HEALTH AND POPULATION SCIENCE RESEARCH CENTERS

The Problem

In the Kingdom of Jordan, several institutions are conducting community health and demographic research with the support of either the United States Agency for International Development (USAID), the United Nations Fund for Population Activities (UNFPA), or the Government of Jordan (GOJ). Although field work typically proceeds on schedule, the processing and analysis of data require considerable external assistance to ensure that work schedules are maintained. This problem persists, despite the installation of large sophisticated computers at research institutions. Because statistical packages are not installed, progress in tabulation and secondary analysis lags. This problem is widely acknowledged to be the result of technical problems with software installation and shortages of skilled computer programming staff. Indeed, these two factors account for some of the difficulties in conducting research. However, there are related issues which should be more widely recognized by institutions concerned with developing research capabilities.

Observations

Where delays are most serious, the preliminary planning of projects is incomplete. The documentation of research plans, with all phases of work specified in a time table, is an invaluable adjunct to data management. Early analytical work, including literature reviews and formulation of hypotheses, provides a conceptual structure for data management. The computer problems in Jordan cannot be solved in isolation from this vital element of survey research.

Data management planning often follows data collection. Serious problems can arise with this approach. When large blocks of unmanaged data are accumulated before computerization begins, coding, computerization, and analysis are delayed while difficulties are resolved.

There is no formal provision for the role of data manager in some of the large-scale studies that are conducted. Data managers can oversee the flow of data from the field, through the checking, editing, and coding process, as well as supervise computer editing, file correction, and file-building. With proper management, data can flow through all stages of development in small lots so that bottlenecks and delays do not arise. The technicians who perform these tasks may be neither computer programmers nor analysts. They should have some understanding of both areas, however, to provide a vital link between computer people and investigators.

Computer programmers typically are assigned to computer facilities and are not administratively accountable to project investigators. In the absence of a data manager, programmers are expected to structure their own instructions for editing and file-building. Because work can involve a succession of programmers, it is possible for redundant, inappropriate, and poorly documented software to be prepared. Because computer personnel are not responsible to project investigators, they cannot be held to deadlines or performance standards.

The availability of computer hardware in Amman is impressive, but systems are often underused. Computer vendors must compete in a limited market: Few systems of any one brand are sold; consequently, hardware sales have, apparently, received more emphasis from vendors than software support services. Software support is expensive in a limited market. Thus, rather limited system options are available on computers, despite the inherent power and flexibility of hardware, perhaps because vendors have not trained users in system operations and procedures. The rather limited interactive capability of the Department of Statistics (DOS) computer, for example, is surprising, given its size and power. A recurrent theme in discussions with investigators is the impression that companies are not providing adequate software support services to their customers.

This problem is exacerbated by the absence of scientists trained in computer management. Computer managers oversee system development, staff development, equipment selection, documentation procedures, job standards, work assignments, and many other issues not directly related to the coding of programs. Well-managed facilities can use the resources available from vendors and fully use the power of equipment. Although highly qualified analyst-programmers may be and have been assigned the task of computer management, they could benefit from special training in the unique features of the hardware installed in their facilities and special technical training in computer management. Short courses in computer management are available at the Asian Institute of Technology (AIT) in Bangkok. Several Asian institutions have found the AIT courses to be useful adjuncts to in-house staff development. Similar courses are, presumably, available in the United States and Europe.

One response to implementation problems has been to set up separate facilities, with the hope that each in-house facility will somehow make computer people more responsive to specific needs. The number of facilities has proliferated, and still more are being planned. If good technical people are available, this is an appropriate strategy. In the absence of a large cadre of competent technicians, however, this strategy may only exacerbate problems. Jordan needs a few well-designed facilities where staff development programs can flourish.

Some facilities, such as that of the DOS, are lacking trained analyst-programmers. Only COBOL programmers are available at the DOS. Although COBOL is useful for many applications, it is not well-suited to statistical computation and special-purpose analytical computing. P-Stat will undoubtedly help, and it is NCR-compatible, but there will always be a need for FORTRAN in any statistical research center.

Recommendations

The consultant's recommendations to resolve problems and improve data analysis in Jordan are summarized below.

1. The GOJ and technical assistance agencies should identify a consultant to review computing problems. The need for such a review has been expressed by everyone consulted and could be formalized by raising the issue with relevant GOJ officials. Dr. Nizamuddin is most eager to assist the GOJ in structuring a request.

2. This proposed consultancy should not be limited to the installation of particular software packages, but should aim also to diagnose computer management problems in Jordan and arrive at a set of recommendations for training the senior staff at key facilities in computer management.

3. Because the technical capabilities of the vendors is apparently weak, the consultant should address the question of how system support services can be improved. Any consultant should visit the vendors as well as the computer facilities to examine this problem.

4. General problems can hamper efforts to assist in the development of solutions to meet the specific needs of population research agencies. The DOS and AID may be interested in population, but the government may be interested in developing central facilities at the National Planning Council (NPC) or elsewhere. Therefore, the DOS request should be specific to a "project" on demographic data analysis. A short-term consultant should examine general problems, such as those noted above, and report policy recommendations to the NPC. A separate and distinct project is needed for population which addresses the problems of two facilities, the University of Jordan and the DOS, over a sustained period.

5. According to representatives of the DOS, the sustained attention of a computer expert is needed for:

   a. Interim computer management at the DOS, if the current manager is sent for training.

   b. Development of a training program in computing in demography at the Department of Population Studies. One course in computer applications, offered once a year, would be a useful addition to the curriculum.

   c. Development of work procedures at the DOS in computing (e.g., work assignments, management control, documentation procedures, testing procedures, etc).

d. Systems programming at the two facilities to improve processing capabilities.

e. Continuous staff development at the DOS in package utilization (P-Stat) and languages such as FORTRAN and BASIC.

f. Evaluation, purchase, and installation of data management and analysis packages.

g. Selection and installation of a mini-computer at the Census Office.

This consultant concurs with all but the last of these recommendations. A mini-computer is needed, as DOS staff explained, because the NCR will not handle statistical packages. An independent consultant should evaluate the capacities of the NCR and the reasons that SPSS will not work there. Although purchase of another computer will compound management problems, it may in fact be necessary, if the NCR is inappropriate for statistical computing. In any case, SPSS is but one of several packages, and alternatives should be considered. P-Stat apparently works on the NCR. It should be installed immediately. This institutional development effort would contribute substantially to research capabilities in Jordan. It would require the sustained commitment of an analyst-programmer with computer management experience.

Conclusions

The problems that lead to delays are more managerial than technical. Software and computers are already available that could greatly expedite analysis. Management problems are of three types: research management by analysts who, in several instances, do not adequately specify the plans for analysis, thus generating inappropriate data sets with complex data processing requirements; underdeveloped computer management; and underdeveloped data management. Large, sophisticated computers have been purchased and installed, but, although programmers are being trained, they are not likely to be able to develop efficient computing centers. Moreover, the basic principles of management of large-scale surveys are often not implemented, so that a given study design and coding often generate large blocks of unprocessed data. Continuous-flow data management can avoid bottlenecks by building edited data files continuously in small lots.

Assistance for institutional development, which is aimed at avoiding data processing delays, should be addressed to the improvement of research planning and documentation; computer systems management; and data management and staff development.

Appendix C

REPORT ON TECHNICAL ASSISTANCE TO THE DEPARTMENT OF COMMUNITY MEDICINE, UNIVERSITY OF JORDAN, ON THE HEALTH STATUS IN THE JORDAN SURVEY OF 1980

Appendix C

REPORT ON TECHNICAL ASSISTANCE TO THE DEPARTMENT OF COMMUNITY MEDICINE, UNIVERSITY OF JORDAN, ON THE HEALTH STATUS IN THE JORDAN SURVEY OF 1980

I. INTRODUCTION

Scope of Work

In 1980, the University of Jordan, Department of Community Medicine, fielded a survey of the health status of the Jordanian population. Funds for this survey were provided by the United States Agency for International Development (USAID). The purpose of the study was two-fold: to provide baseline information on the health status of the Jordanian population and to provide data for the evaluation of the health impact of primary health care (PHC) clinics that have been operating for five years (since December 31, 1979).

USAID requested the American Public Health Association (APHA) to provide technical assistance to the Department of Community Medicine, University of Jordan, for a two-day period in May 1982. Progress on this study has been delayed somewhat; therefore, the purpose of the APHA consultancy was to provide an opportunity for a brief interchange between scientists working on the project and an outsider with experience in managing data so that, collectively, barriers to analysis could be identified and possible remedies discussed. The consultancy involved meetings with investigators, in groups and individually, to review plans, suggest recommendations, and review those recommendations. This report is a summary of the scope of the discussions and of the recommendations completed during this brief consultancy.

In this report, the design of the study--the proposal, the sample, the questionnaire--is discussed first. Next, the central problem with the study--the design of a data structure--is reviewed. Finally, two types of recommendations are highlighted: recommendations for study documentation and planning, and recommendations for further technical assistance.

II. OBSERVATIONS AND FINDINGS

Study Design

A. The Proposal

The documentation for the Baseline Study of the Health Status of Jordan (BSHSJ) consists of a five-page proposal and questionnaires. The proposal describes research methods: sampling (one paragraph, described below); research methods (two paragraphs); implementation (four paragraphs), time schedule (four paragraphs), and budget (one page). Although it affords a general overview of the tasks of the study, the proposal does not address the question of research goals and specific aims.

It is this limitation of the proposal--the absence of a research plan--that accounts for current data management problems. Data management must conform to strict and precise guidelines. These guidelines flow logically from tabulation plans, which, in turn, flow from hypotheses to be tested. The fact that a baseline survey is descriptive does not obviate the need for such research plans.

B. Sampling Design

For this study, two groups were included in the sampling: 10 villages with established health clinics, and 10 villages without health clinics, as of December 31, 1979. Urban areas were excluded from the survey, and all households in a sample village were included. The frame for the samples was arbitrarily limited to villages with a population of fewer than 3,000 households. Thus, the objective of the study was not to produce a representative sample of the rural Jordanian population. Nevertheless, an unspecified stratification procedure was used to disperse the sample villages throughout the Kingdom. Sampling fractions, however, were not designed to select villages in probabilities proportional to size. Thus, some weighting procedure, yet to be devised, will be needed to weight households or individuals according to their respective probabilities of selection. Currently, there is no written description of how this will be done. Existing documents on the sampling procedures indicate, however, that considerable work is needed to review the implications of the sample design for data management and analysis. Weights must be computed and linked to the appropriate unit of analysis before tabulation can begin.

C. Questionnaire

The data collection instruments for this study are in two modules: Form One, which provides general information about the villages, and Form Two, which provides detailed information about the family in five sections. The five sections are described below.

1. The general information module includes a list of members of households, indicators of levels of sanitation (e.g., water resources, presence of animals, and disposal of waste and fuel consumption). The questionnaire elicits recall of illness of each individual. Individuals are listed according to the kind of illness or accident specific to those individuals. If individuals or households did not experience an illness or an accident, names were not recorded. This questioning procedure is followed first for general illnesses or accidents, and then again for specific types of illnesses. Also, a series of questions on demographic events, such as pregnancies or deliveries in the year before the survey, is asked. Finally, there is a section on access to medical facilities, physicians, and paramedics. Following the sequence of questions on accessibility of health services, is a series of questions on health behavior. The questionnaire ascertains, for each illness that occurred in a family, the number of times that the illness was referred to a physician or a clinic and the reasons why or why not government health service facilities were used. The questionnaire also asks about access to non-governmental services and the use of such services. There are series of questions on health beliefs and behavior considered to be correlates of certain health problems, such as smoking, hygienic practices, etc.

2. The second section of the questionnaire is intended to assess the incidence of diseases. For each individual listed in the preceding section as having an illness or accident in the period immediately prior to the survey, there is an open question about the respondent's type of ailment. This question is asked either of the patient himself or an informant, and a code is entered to indicate whether or not the patient answers the question about the episode or the ailment. For each such episode, there is a history of symptoms: how symptoms developed, whether the ailment continued, how long the illness lasted, how the ailment was treated, whether laboratory tests were conducted, and what kinds of tests were conducted, whether a medical examination of the patient was made, whether drugs were prescribed, and what the costs of the drugs were. Patients are asked about their impressions of the quality of services rendered. If the patient was hospitalized, a similar sequence of questions records experiences leading to hospitalization, the duration of care, cost, and the outcome of care. This

section generates a large volume of data because the questions are asked not only of the individuals reported in the first section, but also of every episode of illnesses occurring to those individuals.

3. A questionnaire on pregnancy and delivery records every pregnancy that terminated in the 12 months preceding the survey. Events are recorded even if the respondent is not the person who experienced the pregnancy or the delivery. Questioning begins by asking about pregnancy status and then about medical care during pregnancy. Respondents are asked who attended delivery, whether assistance was sought prior to delivery, and when during the pregnancy assistance was rendered. Finally, there is a series of questions about problems during the pregnancy, complications during delivery, specific treatment for those complications, duration of postpartum amenorrhea, feeding practices, and postnatal care for the infants, including vaccinations and medical examinations.

4. For the fourth section, interviewers are referred to Section One and are asked to record for each death a series of items of information. Interviewers record the date of death, its cause, the duration of illness prior to death, the treatment of the illnesses, and the source of treatment. Finally, there are questions concerning the recording of death, whether the death was officially reported to authorities, the timing of reporting, and whether a death certificate can be produced.

5. The fifth section addresses dietary practices, kinds of food consumed, amounts of food consumed, sources of food, attitudes about food and nutrition, and access to commodity food relief programs. The height, weight, and age of members of the household are taken to assess nutritional status.

Each section of the questionnaire generated a plethora of data, although, as of May 1, 1982, only the nutrition section had been coded, and no data have been punched or edited for processing. The design of the questionnaire reflects a perceived need to record as much information as possible about the health status of the Jordanian population, so that a variety of studies can be undertaken in the future. The scope and contents of those studies have not been documented, however, and the large volume of data that has been collected from the initial study has resulted in serious data management problems. For these reasons, USAID requested the APHA consultancy, with the concurrence of the Department of Community Medicine, to clarify current data management problems, to discuss those problems with the investigators, and to assess whether further technical assistance will be needed in the future. Thus, an evaluation of the questionnaire is beyond the scope of this consultancy. It is fitting, however, to review briefly how questionnaire design contributes to current data management problems.

The most important problem is that levels of analysis are mixed on a given module: A household questionnaire addresses questions about individual health problems. Any question on a given individual in one section of the questionnaire cannot easily be associated with other information about the individual elsewhere in the questionnaire.

The coding must provide for a variable length of code sequences: Different number of individuals experience illnesses from household to household. Thus, tabulation and editing will be complicated by the fact that the amount of information coded for a given unit of variation is in itself variable.

Editing will be complicated by ambiguities about the person responding to the questions. The capacity to check on the logical integrity of the data is impaired when one does not know who the respondent is and what attributes that individual has, or whether the respondent changes from question to question or module to module.

Finally, the structure of the questionnaire is inconsistent with the plans for analysis. This structural problem required considerable discussion with scientists involved in the study and merits review in this report.

D. Data Structure

As noted above, the investigators have structured the data in five modules. The plans for analysis, as verbally presented by the investigators encompass three levels of analysis: correlates of village, family, and individual characteristics of disease episodes or mortality risks. Current data management procedures, however, do not structure the data according to this plan of analysis. Instead, coding plans structure data as they are specified in the questionnaire forms. Thus, there is one section of coding for the village module, and five separate coding plans for Questionnaire Form Two, the most ambitious of which is Section 2, containing questions on the health status of the family. To understand the more serious difficulties with the coding plan, consider the following, rather trivial, model for the risk of death of a child:

$$\Omega = \lambda + \lambda_1 N + \lambda_2 A + \lambda_3 H \qquad .....(1)$$

where Ω = the odds of death (or whatever), 1 = death, 0 = surviving:

λ = an intercept

λ_1 = the effect of nutritional status

λ_2 = the effect of age

λ_3 = the effect of health facilities,

and N, A, and H are variables representing nutritional status, age, and access to facilities, respectively.

Each variable is on a different module of the questionnaire, each with a separate code plan and code layout. Clearly, data must be linked; so, the single merged file exists wherein the individual is the basic unit of analysis and characteristics of individuals, families, and villages are coded together to permit analysis on the computer. The linkage of record files with the requisite dependent and independent variables in a common unit of analysis is not developed in the current coding plan. Alternatively, families could be the unit of analysis, with rates of death or illness serving as dependent variables. This consultant did not address the issue of which analytical plan was best; he instead focused on the logical link between alternative plans for analysis and the requisite data structures for each plan.

The model in (1) does not need the estimates, but Ω will have to exist in some form so that tables can be constructed. To conduct the analysis and fulfill the goals presented by the investigators, there must be a string of dependent variables for each individual representing mortality or morbidity for particular individuals or households. The code plans specify a list of the kinds of illnesses or accidents that might occur in a household, and names are recorded in addition to the illnesses or events. If no illness was present, no name is recorded. This approach eventually will create difficulties for analysts. The single most difficult problem in building a data structure is that in no case is the dependent variable available for analysis. This is because only events that occurred are recorded and there is no variable in a fixed field to indicate whether or not a particular illness or condition occurred.

In summary, there are two fundamental problems with the data set. One, the file structure is not suitable for analysis, because units of analysis will not contain the requisite data from different modules; and two, the dependent variables are not coded as variables. Information is present, but not in a form suitable for tabulation. Consequently, substantial data management will be required after coding and data entry have been completed.

III. RECOMMENDATIONS

The recommendations for this project are divided into two categories: recommendations to document data management procedures and recommendations to secure further technical assistance.

Recommendations for Data Management Documentation

In conducting a survey of the size and scope of the Baseline Study of the Health Status of Jordan, it is important to have a series of documents specifying the layout of data, the plans, and the steps in data management. Some possible documents were discussed in considerable detail with the investigators during the brief consultancy. The following presentation is a paraphrase of written examples of the documentation left with the investigators.

A. Documentation Note 1: Preliminary Research Plans

In planning the layout of the data, it is essential that each investigator attempt to document research that is planned for the first several months after the data have been prepared. To illustrate this point, two possible research reports may be considered. The topics are hypothetical, but they illustrate the importance of documenting data management plans.

1. Topic 1

The purpose of this study is to investigate the correlates of child mortality and morbidity in Jordan. It is assumed that the background characteristics of individual children, households, and villages affect the incidence of specific childhood diseases.

2. Topic 2

This study intends to investigate the hypothesis that families in Jordan differ with respect to risk of illness, and that the composition of households, the characteristics of household members, and the characteristics of villages where members reside determine the kinds of illnesses that arise.

In a series of practical examples, it was demonstrated that Topic 1 and Topic 2 have quite different implications for the way that data may

be coded and managed. For Topic 1, a separate record would be needed for each individual, with variables for each potential source of illness. For Topic 2, however, the procedure would be to study the incidence of illness in families. This procedure is being used in the coding plan for the study described in this report. Yet, the investigators have professed an interest in Topic 1-type analysis. For this reason, the consultant aimed to clarify units of analysis and their structural implications. Once plans of analysis have been specified in detail, it will be clear to the investigators that current data management plans are inconsistent with the plans of the analysis. The following illustrative example is extracted from notes left with the investigators:

"To illustrate this, consider a simple and rather basic "Topic 1" table: the frequency of illness of any kind against age of person in the year before the survey.

"The dummy table looks like so:

The Frequency of Complaints of Illness in the Past 12 Months by Age of Household Member

Experience of Illness	0 - 4	5 - 14	15 - 34	35 - 54	55+	Total
Yes (Percent)	n_1 (%)	n_2 (%)	n_3 (%)	n_4 (%)	n_5 (%)	n
No (Percent)	$N_1 - n_1$ (%)	$N_2 - n_2$ (%)	$N_3 - n_3$ (%)	$N_4 - n_4$ (%)	$N_5 - n_5$ (%)	$N - n$ (%)
Total	N_1 (100%)	N_2 (100%)	N_3 (100%)	N_4 (100%)	N_5 (100%)	N (100%)

"To compile this table, the following logical steps are required, assuming No = 0 and Yes = 1 for our dependent variable Y:

"1) Go to Form 3, Section 1, Q205, and access line number 1. Label it I_1.

"2) Now go to Q205 and access age for that individual. Label it A_1.

"3) Skip to Q2221 and test for illness in the two weeks prior to the survey. If no illness occurred, skip to logical step 5. If an illness to I_1 occurred, go to Q222 and conduct logical step 4.

"4) Now complete the following:

Is line 1 = I_1 in Q205? If yes, go to final step 7: $Y_1 = 1$.

If no, test line 2. Is line 2 = I_1 in Q205? If yes, go to final step 7, and code $Y_1 = 1$.

Now proceed to line 3, line 4, etc., through 10 lines, searching for $Y_1 = 1$, whenever line 1 from Q205 appears. Whenever $Y_1 = 1$, skip to step 7.

After doing so for 10 lines, proceed to step 5 if I_1 has not been found in Q222.

"5) At this point it is known that for individual I_1 no illness occurred in the two-week period prior to the survey. But risks in the period 3-52 weeks prior to the interview are not known. Skip to Q227. Is Q227 Yes = 1, or No = 2? If no, then $Y_1 = 0$, and skip to final step.

"6) Now it is known that someone got sick, but it is necessary to ascertain if it is individual I_1. Therefore, the logical steps 1 in (4) above must be repeated for Q228:

Is line 1 Q228 = I_1? If yes, $Y_1 = 1$, and go to final step 7. If no, test the next line number. This must be repeated as in step 4.

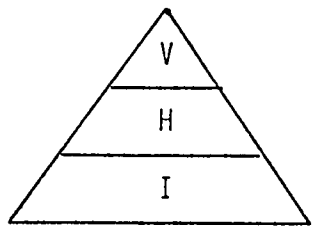
"7) At the final step the 10 lines of Q228 have been scanned. If $Y_1 = 1$ at this point, then I_1 never got sick and $Y_1 = 0$.

"This completes one family. As many as 26 logical steps are involved in tallying an individual case for the dummy table. Clearly, the information is there, but the data management task is complex."

B. Documentation Note 3: Structure of Data Files

Given the analytical plans, it is appropriate to document overall data structure: what is varying for the purpose of analysis. From study documents, it appears that there are three levels of variation: villages, households, and individuals. Investigators should show how they will produce the data for these three levels and how they can link them together

into a "nested data set" that resembles a pyramid; for example:



Village

Household

Individual

When analysis is in progress, the records might look like this:

Case 1	I_{11}	H_1	V_1
Case 2	I_{21}	H_1	V_1
Case 3	I_{12}	H_2	V_1
Case 4	I_{22}	H_2	V_1
-	-	-	-
-	-	-	-
-	-	-	-
			V_2
			V_2

Or, alternatively, levels of aggregation might be separated. In any case, what is done in structuring the data will depend upon the analytical plans.

The consultant's advice for Documentation Note 2 was to go through all the questionnaires and compile three lists of the variables, according to level, and then to clarify the overall structural requirements of the research plans.

C. Documentation Notes 3-8: Detailed Coding Instructions for Data Sets as Currently Structured

There should be one set of coding instructions for each of the six data modules. Coding should specify <u>on each card</u> the identification

sequence, and it should be clear from the written instructions that all the cards from the project can be placed in a computer in any scrambled order and still be sorted by village, family, and module. No coding plan has been documented, yet a substantial amount of coding has been completed. The existing code structure should be documented in detail, column by column, record by record, for the six modules. The development of a final record format that is consistent with the plans of analysis (Note 1) and overall data design (Note 2) will require thorough documentation of what has been done or will be done with the data. The consultant recommended coding the following data:

1. Identification Sequence

   Every record must have an identification sequence so that records can be sorted by village, family, module type, and card number. The layout for every card format should provide for sufficient columns to identify the particular card that is being coded.

2. Sequential Code Format

   After the identification sequence, it is important to specify, in detail, all the information embodied in each column of coding. Typically, this involves printed columns with the following items:

   a. Sequential Variable Number. It is appropriate to assign for each coded field, V0001, V0002, ...etc., so that in the next documentation note the computer logic can be specified to the programmers using the variables.

   b. Variable Label. Each variable should have a unique five-character (or fewer characters) name for SPSS. These names will appear in an SPSS dictionary and will be printed as SPSS output. Several examples are:

AGE	Age of Respondent
RELAT	Relation to Head
OCCHD	Occupation of Head.

   c. Variable Description. The full meaning of the variable name should be specified in written words; for example, "age of respondents," "relation to head," etc.

   d. Variable Location. The variable location specifies the range for the card and column.

e. Code Classes. For each possible code, it is necessary to list for each digit the meaning of each code and the instructions for coders.

D. Documentation Notes 9-14: Computer Editing Instructions

Each module has logical integrity. It would be appropriate, therefore, to require someone to examine each variable in each module and specify its range and every conceivable logical check. In this way, a programmer can be shown how every variable relates to every other variable. Careful and meticulous attention to every logical detail is vital to converting data from its current format into something that can be analyzed. Logical flaws in the data will have to be corrected before the steps in Notes 15-17 can be completed.

E. Documentation Notes 15-17: Final Data Format for Village, Family, and Individual Level Data

In Documentation Note 2, the general appearance of the data is described in that the pyramid and the general contents of each level of data are described. The series of documentation notes should specify the exact format of data in three final code layouts. The document should resemble the coding plan described above, except that the instructions would be logical instructions to the programmer.

Three files will emerge from this exercise: village (Note 15), family (Note 16), and individual (Note 17). Note 17 will be difficult to write, because of the complex logic involved. What will be required is the written logic that generates a row file of individuals, with column after column of data about each individual.

Recommendations for Technical Assistance

After the basic documentation for the study has been completed by the investigators, several rather sophisticated computer programs will have to be written to edit the data, link and build files, and conduct analyses. Technical assistance in two areas, possibly from Jordanian institutions and individuals, will be needed to ensure that the study is analyzed within a year: data management assistance and the assistance of an analyst-programmer.

A. Data Management

Data currently are coded on precoded forms included in the questionnaire for this purpose. Three coders are working on the data, all of whom are professionals with an interest in analysis. Large batches of data are backlogged; priority has been given to coding the nutrition module. By assigning priority to a module, the investigators can reduce the time required before data analysis can begin--an appropriate strategy in light of the difficulties that will be experienced in processing other modules.

The current coding operation should be an interim arrangement, however, designed to get a piece of data into the analytical phase. To complete the study, a more intensive coding operation that involves simultaneous supervision of coding and editing by an experienced data manager will be needed. Coding procedures should be modified so that several stages of data management (as outlined in the recommendations in the documentation notes) are under way with small blocks of data. In this way, specific editing and computer bottlenecks will become evident before large blocks of data enter a particular stage of development.

It is recommended that a professional data management specialist be recruited who can undertake the following tasks:

- Consultation with scientists to develop documentation, if necessary.
- Preparation of editing instructions, in consultation with investigators.
- Development of work flow, through computer editing and supervision of computer editing, to the checking and correction of data.
- Recruitment and training of a small coding staff to expedite completion of coding.
- Orientation and direction of a computer programmer for the project.

B. Software Development

Investigators repeatedly expressed an interest in SPSS, and noted the university's progress in installing SPSS on the IBM S4331 that is available to them. It is important to note, however, that in this study, SPSS will be of greater use during analysis than during the preparation of data. Several sophisticated special-purpose programs are needed for editing and file-building. It is urgent that a person be identified who is funded by

the project and reports to its investigators. A system for computer editing and file-building must be developed. Because of the complexity of this study and the problems with data structure, an analyst-programmer will be required for approximately six full months.

IV. CONCLUSIONS

The Baseline Study of the Health Status of Jordan is a large study requiring the completion of major data management tasks. Whether the study can achieve its objectives or not will depend upon the care with which the investigators clarify their analytical plans and structure the data so that they are consistent with those plans. During the consultancy, a recurrent theme was the need for research planning to overcome data-processing problems: Data will never be ready for analysis until analytical plans have been formulated that can guide data managers and programmers. The investigators should, therefore, not expect a programmer to be added to the staff, nor should they expect SPSS to be installed to solve the problems that are ahead.

There is some liklihood that a request will be made for support for computer assistance for this project, and that such a request will merit careful consideration from donors. If research plans are developed and documentation is improved, the data could become a useful research tool for understanding the health problems of rural Jordan.