BIBLIOGRAPHIC DATA SHEET	1. CONTROL NUMBER PN-AAH-268	2. SUBJECT CLASSIFICATION (695) PA00-0000-G200

3. TITLE AND SUBTITLE (240)

Some results and problems on the estimation of vital rates in a rural African setting via multiple mrthods

4. PERSONAL AUTHORS (100)

Kpedekpo, G. M. K.; Wurapa, F. K.; Lourie, I. M.; Nuemann, A. K.; Belcher, D. W.

5. CORPORATE AUTHORS (101)

Calif. Univ., Los Angeles. School of Public Health.; Ghana Medical School.

6. DOCUMENT DATE (110)	7. NUMBER OF PAGES (120)	8. ARC NUMBER (170)
1975	33p.	GH312.C153a

9. REFERENCE ORGANIZATION (130)

Calif.--LA

10. SUPPLEMENTARY NOTES (500)

(Danfa Comprehensive Rural Health and Family Planning project, Ghana, mono. ser. no 4)

11. ABSTRACT (950)

12. DESCRIPTORS (920)

Rural areas
Ghana
Estimating
Vital statistics
Demographic surveys
Health data collection
Statistics
Family planning
Health services
Population policies
Public health

13. PROJECT NUMBER (150)

641005500

14. CONTRACT NO.(140)

AID/CM/afr-IDA-73-14

15. CONTRACT TYPE (140)

16. TYPE OF DOCUMENT (160)

AID 590-7 (10-79)

SOME RESULTS AND PROBLEMS ON THE ESTIMATION OF VITAL RATES IN A RURAL AFRICAN SETTING VIA MULTIPLE METHODS

Monograph Series, Number 4

DANFA COMPREHENSIVE
RURAL HEALTH AND FAMILY PLANNING
PROJECT, GHANA

G. M. K. Kpedekpo, Ph.D.
F. K. Wurapa, M.D., M.P.H.
I. M. Lourie, M.D., M.P.H.
A. K. Neumann, M.D., M.P.H., F.A.B.P.M.
D. W. Beicher, M.D., F.A.B.I.M.

The data appearing in this paper are preliminary. The paper may not be quoted in whole or in part or otherwise used except with the written permission of the project co-directors.

March 1975

FOREWORD

The Danfa Comprehensive Rural Health and Family Planning Project is a teaching, research, demonstration and service programme initially developed by the Department of Preventive and Social Medicine of the Ghana Medical School in 1965 and implemented under contractual agreement with the School of Public Health at the University of California, Los Angeles, and the Agency for International Development in 1970. The eight year project is closely associated with other Ghanaian agencies and institutions such as the Ministries of Health, Social Welfare, and Finance, the Ghana National Family Planning Programme, and other University of Ghana departments.

The Ghanaian institutions have primary authority and responsibility for project administration. Ghanaian staff are hired by the institutions at prevailing terms of employment. All service costs are borne by the Government of Ghana as well as some of the training and evaluative costs. UCLA/AID are responsible for most of the research costs.

The project includes a longitudinal study of approximately 50,000 rural inhabitants, 10-50 miles from Accra. The project area has been divided into four sub-areas to investigate alternative methods of delivering health and family planning services, summarised in the chart below.

	Service Provided			
Area	Comprehensive Health Care	Health Education	Family Planning	Standard Health Services
I	Yes	Yes	Yes	Yes (equivalent)
II	No	Yes	Yes	Yes
III	No	No	Yes	Yes
IV	No	No	No	Yes

One of the principal goals of the project is to provide information about project activities and findings to concerned agencies on an ongoing basis to assist in health planning. A project monograph series has been developed to aid in this information sharing process, and we hope that the data contained in these papers, although often preliminary in nature, will prove useful to those involved in the project and to others interested in obtaining current information on rural health projects in developing countries.

Alfred K. Neumann
Associate Professor
UCLA School of Public Health

For
S.R.A. Dodu
Dean, University of Ghana
Medical School

Co-Directors
Danfa Comprehensive Rural Health
and Family Planning Project

The Danfa Project is supported in part by USAID Grant: OCOP/USAID-CM/afr-IDA-73-14.

This paper is one of a series of vital events and demography articles which report on data obtained from vital events registration and baseline demographic surveys conducted for use in the Danfa Project and by interested agencies and institutions in Ghana. The data presented here are, however, preliminary, and subject to modification after further analysis. Updated data will be issued as they become available.

Acknowledgement

The authors gratefully acknowledge the services of Dr. A.E. Okorafor and the supervisors at the Legon Data Processing Center as well as the diligent efforts of the field research assistants (interviewers) at the Legon Center.

Some Results and Problems on the Estimation of Vital Rates in a Rural African Setting via Multiple Methods

1. Introduction

Many countries in Africa have, in recent years, recognised the need to adopt population policies designed to keep in check accelerating growth rates. The formulation of population policies and the assessment of the effect of such policies require precise, categorised and up-to-date information on the dynamics of population change. Effective management of a public health programme requires the compilation of detailed demographic information on age, sex, mortality, cause of death, and so on. Such data cannot reliably be obtained from survey alone, even when repeated at regular intervals. Current reliance on survey and census data alone has proven insufficient for the demands of the present and future. These serve only as a stop-gap measure until such data may be combined with data obtained from a continuous vital registration system.

A system based on a combination of census and conventional registration data is difficult with regard to costs, organisational constraints, accuracy and coverage, even when fully developed. Hence, Africa and numerous other developing countries have, for the past twenty or more years, increasingly directed their efforts toward the development of alternative sources of demographic information and of estimates of population growth. Every country presents it special and different problems and possibilities in regard to data collection, necessitating experimentation and innovation

for the development of appropriate systems.

Such experiments are essentially prospective in nature and consist of ongoing vital events registration and a retrospective procedure covering the past reproductive history of subjects, applied to the same population. These experiments afford great flexibility and allow for almost any item of a demographic nature to be included on an ad-hoc or more continuing basis. Through reconciliation of the two procedures by means of matches, adjusted vital event rates are obtained which are more accurate than those obtained when either procedure is used alone. These methods, are novel by virtue of the collection procedures and analytic techniques utilized. Assessment of the success of the new methods cannot be based on the collection procedures or the analytic contributions alone. Procedural and cost factors both must be accounted for in the difficult task of assessing the value of the methods for general and specific purposes.

The follwoing are the results of recent experiments in the collection and analysis of data on births and deaths by use of multi-approach methods, specifically the estimation of vital rates by the well-known Chandrasekhar-Deming method. In this description, the experiences of the Danfa Project will be central and the main problems in the use of the method will be indicated as they apply to the Project. The paper thus summarises achievements, failures and frustrations and contributes to the body of knowledge on data collection and analysis of vital registration data, using a multiple methods system in a rural African setting. Modifications and developments of existing practices are also considered as well as recommendations based on our experience.

2. Data Collection Methods and Procedures

The method employed in the collection of data on births and deaths is the follow-up survey in which required visits within specific periods are made. Previous to the survey, a baseline enumeration was undertaken to gather basic information on age, sex, family relationship, education, religion, ethnic group and pregnancy status. Follow-up is conducted at six-month intervals to cover recent births and deaths, including infant deaths. These semi-annual and annual resurveys were designed to increase the scope for matching and the accuracy of the data on births, deaths and migration. These re-checks and resurveys were independent of the vital registration records, consisting of births and deaths as they occur in their respective village clusters, on a continuing basis.

The six-monthly and twelve-monthly resurveys are expected to give accurate results for data on births, deaths and migration, thus accounting for changes in household composition. This procedure should give improved data on current fertility as errors of dating are nearly eliminated, moreover, surveys include a question on pregnancy, rendering probing in follow-up rounds more effective.

In the case of birth data, one may start by asking about births to women listed, but eventually one must also ask about births to women who were not listed or who may have died since the previous round. In addition, in societies in which women tend to bear children in the homes of their mothers, questions such as the following were used as probing questions : "Since my last visit did a daughter or anyone else visit you and have a baby in the household?" In compiling death records, the multi-round follow-up survey is advantageous since enumerators can enquire about the survival of named

individuals who were in the area the previous time the enumerator visited the household; it further enquires about other deaths (e.g. of visitors) in the household since the previous visit. The follow-up method also obtains important migration information on the movements of persons into and out of the area and between and within areas between survey rounds.

The development of a continuous registration system requires consideration of several factors. One of these considerations is the base populations. There is an important distinction between enumerating vital events which occur over time and enumerating the base population which is counted at one time, in that de facto and de jure concepts do not mean the same thing* when referring to vital events and when referring to people.

For example, the residents at the time of the survey may not be identical to the residents for whom vital events are registered during the course of the year. Household surveys usually use a de facto concept for counting vital events. An additional and basic constraint arises in the case of most PGE (Population Growth Estimation) studies: both collection systems must use the same rule for counting vital events, i.e., births and deaths.

A central issue in choosing between the de facto and the de jure approach is the relative difficulty of obtaining reports of certain events. The de facto count must include events for short-term visitors in the sample area, whereas the de jure count must obtain reports of events occuring to usual residents who are temporarily away from the area. In deciding between the

---

*De facto vital events are those which occur among all people within a defined area at a given time (including visitors). De jure vital events occur among usual residents of an area (including those who may be away for brief periods of time).

two approaches, a primary concern is the potential of each for minimizing bias due to double counting, omissions due to missing events and costs accrued in making the chosen concept operational. Training of registrars and the ability of respondents to answer the questions are additional factors which influence the effectiveness of either approach.

Migration can be a serious problem in a project design of this nature. Possible false entries in one of the two reporting systems may appear due to in-migration to or out-migration from the area covered. The effects of temporary migration are often a strong influence on the phenomena under observation. Deaths and births in the sample population may commonly occur in a hospital which attracts outsiders. Similarly, many women return to the home of their parents for the birth of a baby. Accordingly, if survey and registration are based on a de facto definition of the population (which would be the simplest solution from an administrative viewpoint) there is a definite risk that the results will show false discrepancies and the total deaths and births will be over or underestimated depending upon the location of health facilities in the area. This risk can be reduced or eliminated if the same rules are applied to both collection systems. By adopting a de jure concept the problem would be solved but only at the cost of following up residents moving out of the area and keeping track of events affecting temporary residents or visitors. The technical difficulties involved in each solution are likely to be formidable. In view of these facts, a comprehensive registration system was adopted with de jure and de facto counting rules integrated into the system.

Whatever rule is used, it must account for three broad classes of events, and these three broad classes must be accounted for in the re-surveys and registration of vital events. They are as follows:

i) those that occur in the sample area to usual residents;

ii) those that occur in the sample to usual residents when they are away from the area;

iii) those that occur to visitors to the sample area.

Table 1A describes the counting rules for the classes of vital events covered.

Table 1A

To Show the Various Counting Rules in the Collection of Vital Events in the Project Areas

Description of Situation	De Facto	De Jure	Comprehensive
Status at time event occurred			
1. *Mother or person reported dead is/ was a usual resident and physically present in sample area;	Count	Count	Count
2. Mother or person reported dead is/ was a usual resident and temporarily away from the sample area;	Ignore	Count	Count
3. Mother or person reported dead is/ was a usual resident elsewhere and is a temporary visitor to the sample area.	Count	Ignore	Count

*Regardless of the legal requirements, it is easier to classify births in terms of the mother's rather than the father's residence.

The counting rules adopted in Table 1A are the basis of the classification of the residential status of each event according to Table 1B. The details for different types of vital events which should be reported from each source and their definitions can be shown as follows:

Table 1B

Classification of Vital

Events by Resident Status

Residential Status	Registrar	Survey
URP 1	URP 1	URP 1
URP 2	URP 2	URP 2
URA 1	URA 1	URA 1
URA 2	URA 2	URA 2
V 1	V 1	V 1
V 2	V 2	V 2

URP 1 Live births or deaths to normal residents who were present in Area at the initial reference date and found present or deceased at the time of the subsequent visit. (Birth or death took place in the area).

URP 2 Live births or deaths to normal residents in the area who outmigrated during the period between visits, and not found at the time of visit as they were not present in the household or the whole household had migrated.

URA 1 Live births to normal resident occurred outside the area, resident has now returned to the household and present at the time of visit; or normal residents who died during the period outside the village.

URA 2 Live births to the normal resident occurred outside the area, resident has not returned to the household or has migrated out.

V 1 Live births or deaths occurred in the area to the visitor who has visited a household and now returned to the usual place of residence or still present in the household.

V 2 Live births or deaths to the visitor who had come to maternity home/institution and returned to the usual place of residence after delivery.

The various types of vital events enumeration can be shown as follows:

1) comprehensive vital events: URP 1 + URP 2 + URA 1 + URA 2 + V 1 + V 2
2) de facto events: URP 1 + URP 2 + V 1 + V 2;
3) de jure events: URP 1 + URP 2 + URA 1 + URA 2;
4) fixed events: URP 1 + URA 1.

The fixed events are introduced to eliminate the effects of migration from the estimates and to concentrate only on persons and events which are seen and are present in the project areas at each successive round. The extent to which this latter approach has been successful would be determined when the estimates of the rates are made by way of the different methods.

3. Matching of the Registration Data and Survey Data: Problems and Experiences

Matching of vital events reported by the two different procedures (registration and survey) was difficult and time consuming but was necessary for the evaluation of both the registration and survey records. The main steps adopted consisted of mechanical operations, several manual operations and a field check of partially matched or non-matched events. Problems arose due to differences in receiving and recording information on names, and in the large number of items on the reports that are to be compared and verified. Problems involving memory decay, or lapse over time, inconsistency in the members of the household who supplied the information, in mobility of individuals and households over a period of time in the project areas, and other factors all suggested that some uniform method should be developed so that matching could be done with greater accuracy.

The usual procedure in our matching technique was to compare events from the registration with each of the events from the survey. Thus each comparison is obtained by giving each characteristic a comparison of one of the following values:

i) reports agree on each item;

ii) reports contradict each other;

iii) no entry for the characteristic for one or both groups.

Agreement and contradiction are defined with reference to tolerance limits set up to allow for common errors and distributions in the reporting characteristics. After the initial office matching and categorization of events into those which are clearly matched, probably matched, probably unmatched, or clearly unmatched, a follow-up team was sent to the field for re-verification of all the doubtful cases.

The enumerators were particularly requested to check on the unmatched cases for geographically out-of-scope errors and to classify them on the basis of evidence such as: the location of household near the boundary; disagreement on exact date of birth/death; date of birth/death near the boundary of the reference period; date of birth/death unknown; changes in the household composition between rounds. Geographical out-of-scope errors may largely be a function of boundary, definition of areas which may arise from difficulties with area boundaries, and problems of associating the birth or death uniquely to a particular area; the follow-up teams were instructed to look out for these types of out-of-scope cases and to identify them accordingly.

One of the major problems in matching vital events reports was the selection of matching criteria. Too stringent a criteria may cause the same event to be counted as two events and too loose a criteria may result in different events being considered as the same event. However, based on the outcome of other matching procedures and on our own pilot experiments, a number of variables were selected for matching our vital events: the house number; the name of the child/deceased; the sex of the child/deceased; date

of birth/death; place of birth/death (and residential status); name of the mother for births; and the age at death for the deceased.

While the name of the child is one of the criteria used for the matching procedure and an important matching criteria on discrimination, the spelling and the system of naming among the different cultural or ethnic groups in the project areas lead to difficulties in reconciling the same events. In most cases, the names reported to the registrar were different from the names reported to the enumeration team. The actual names of babies were not available in a substantial number of birth reports by vital registration assistants, probably due to cultural variations in the naming of babies after birth.

The house number is another criterion considered important in matching information. A unique feature of the Danfa Project is the careful mapping of all villages and dwellings carried out before the field survey and the registration started. Discrepancies or mismatches in house numbers were common, however, their verification was relatively easy. In most cases these mismatches were the result of cultural variations. For example, in some societies pregnant women may leave their husband's households where they were enumerated to deliver their babies in their mother's or grandmother's houses.

Another variable which was difficult to manage and one which often leads to mismatches, is the date of birth. Here a tolerance limit of plus or minus one month was allowed for accepting the date as a match. Dating of events in a non-numerated and illiterate society accounts for this particular difficulty and it is one which has been noted in many surveys in Africa. The sex of the child was well recorded and easy to match in almost all cases.

Immediately after each enumeration, the lists of pregnant women were compiled and distributed to both the registrars and the survey team. The

registration assistants were requested to follow up these pregnant women for the outcome of their pregnancies. Recording the pregnancies of women improves the accuracy of the reporting of events, particularly infant deaths. Both sources of data collection have access to the information but one difficulty arose in the follow-up exercise which was brought about by mobility in the project area. Enumerators were not able to obtain sufficient information about children of out-migrated parents: the name; the date of birth; sex; place of birth and whether the children born to those out-migrated women were alive or deceased. Despite intensive field verification, it was not possible to account for the outcome of some of these pregnancies and these omissions constitute some element in under-reporting. In general, the pregnancy data was found useful (the full use of the pregnancy data will be discussed later).

During the matching exercise we were aware that one of the basic assumptions in the application of the Chandrasekhar-Deming formula is the statistical independence of the two systems of data collection. Lack of independence may arise because of collusion between the registrars and the enumerators and the possibility that the same type of event might be missed by both systems. With a combination of rigorous field controls and weekly collection of the vital events data from the registration assistants, we were able to ensure some independence of the two systems. Other problems encountered deal with the manual procedures adopted for the matching. As previously stated, the manual matching operations were time-consuming, contributing to the delay in publishing the results. Other less time-consuming systems should therefore be explored. One such possibility is to do all initial matching by computer and to do the field verification by the enumerators. Because of the record system used in ths project (see Section 2 above), this approach seems feasible and worth attempting.

4. Methodology of the Estimation of Vital Rates by C-D Method

The estimation procedure was based on the theoretical work of Chandrasekhar and Deming (1949) which is fully discussed in the literature. The essential features of this procedure is collection of data by two independent systems: the registration system and the periodic household surveys. The vital events collected by the two systems are matched so that the following categories can be obtained:

C The number of events recorded by both systems (registration and survey);

N_1 The number of events recorded by survey only;

N_2 The number of events recorded by registration only.

From these threee categories, estimates of events missed by both systems (under the assumption of statistical independence between the two systems) were derived using the Chandrasekhar-Deming formula which can be given as follows:

$$N = C + N_1 + N_2 + X$$

where X is the estimated number of missing events and can be estimated as $\frac{N_1 N_2}{C}$

this leads to $N + C + N_1 + N_2 + \frac{N_1 N_2}{C} = \frac{(C + N_1)\ (C + N_2)}{C} = \frac{RS}{C}$ (1)

where $S = C + N_1$ and $R = C + N_2$ (2)

with the standard error of N as $\frac{\sqrt{N} q_1\ q_2}{P_1\ P_2}$ (3)

where $P_1 = \frac{C + N_1}{N}$ and $P_2 = \frac{C + N_2}{N}$ (4)

and $P_1 + Q_1 = 1,\ P_2 + Q_2 = 1$ (5)

On rearranging the formula $N = \frac{RS}{C}$, the formula estimates the completeness of the coverage of the registration system as the match rate of the survey, and estimates the completeness of the coverage of the survey as the match rate of the registration. This leads to the following equations for the completeness of coverage rates shown as:

$$\frac{R}{N} = \frac{C}{S} \quad \text{for the completeness of registration;} \qquad (6)$$

$$\frac{S}{N} = \frac{C}{R} \quad \text{for the completeness of survey.} \qquad (7)$$

The major theoretical features of the method are few, simple in conception, and difficult to apply but valuable owing to their unique self-checking features. The application of the procedure to our data produces the results shown in Tables 2 and 3.

5. Results of Coverage Estimates for Births and Deaths

Estimates for completeness of coverage by the two collection systems are shown in Table 2 and 3 for births and for deaths. In general, it is clear that the estimated coverage of both live births and deaths are lower in the first twelve months of operation than in the second twelve months. The variation in the coverage estimates from one area to another probably reflect differences in sources of bias operating in the different project areas as well as differences in coverage. In most cases, lack of independence can be expected to over-estimate coverage, and spurious reports to under-estimate coverage. It is also clear that the estimated coverage for live births is typically higher than that for deaths in all four areas. Area I seems to be more efficient in the collection of births and deaths by the registration method than the other three areas.

The standard errors of the estimates were also calculated. The magnitude of the standard errors for estimated total births and deaths calls for some comment. In general, the larger the size of the standard error of an estimate, the less stable or reliable the estimate itself is. It seems that, from the comparative sizes of the standard error for births in the first year of the project operation, the Area III estimate of total births is the least reliable, followed by Area II, and IV in that order. In the second year of operation, there is an overall general improvement in the sense that there

is some reduction in the sizes of the standard errors, with the exception of Area I. Area III is still the highest and therefore the estimate itself is the least reliable. This is due to the relatively small number of common events recorded in Area III by both the registrars and the enumerators. The sizes of the standard errors for deaths follow essentially the same pattern as those of births, although for Areas II and III they are larger, during the two periods. The reason here, as with births, is the low percentage of events caught by both systems during the two successive years of operation. A more detailed examination of the coverage rates for live births and deaths is given in a separate discussion, but the pattern of the standard errors indicates that some particular bias may be operating in Area III and in the other areas to a lesser extent.

Registration coverage or percentage efficiency for live births for the first twelve months of operation varies from area to area, showing 53.2% coverage for survey only, 59.4% for registration only and 81.0% for both methods for Area I. The corresponding figures for Areas II, III and IV are shown in Table 2. For the second twelve months of the operation it is clear that coverage has improved considerably in all areas although the estimated absolute total remains approximately at the same level as the previous twelve months, thus for Area I, the total percentage coverage by survey method only is 78.5%, by registration method only, 65.7% and by the two methods combined, 92.6%. The figures for Areas II, III and IV are shown in Table 2.

The survey team covered 47.9% of the deaths in the first period of operation (12 months) and the registration covered 69.1%. The total combined method covered 83.9%. In Area II the percentage coverage for survey, registration and combined methods are 32.4%, 24.5% and 48.9% respectively. During the second period of operation, the coverage from survey method

Table 2

Percentage Distribution of Matching Results for the 1st and 2nd 12 months by Type of Enumeration - Births

Area	Method	Common Event C	Survey Only N_1	Registration Only N_2	Estimated by C-D Method E	$C + N_1 + N_2 + E = N$ N %	N (Abs.)	Standard Error of N	Total Coverage by Survey $C + N_1$	Total Coverage by Registration $C + N_2$	Total Coverage by both $C + N_1 + N_2$
1st 12 months											
I	Comprehensive	31.6	21.6	27.8	19.0	100.0	611	19.2	53.2	59.4	81.0
	De Facto	35.2	18.6	30.3	15.9	100.0	469	8.7	53.8	65.5	84.1
	De Jure	31.7	21.5	27.9	18.9	100.0	595	18.8	53.2	59.6	81.1
	Fixed	35.3	19.6	29.0	16.1	100.0	504	15.2	54.9	64.3	83.9
II	Comprehensive	22.0	33.8	17.4	26.7	100.0	672	28.5	55.8	39.4	73.2
	De Facto	24.2	26.7	23.3	25.8	100.0	464	22.2	50.9	47.5	74.2
	De Jure	21.9	31.0	19.5	27.6	100.0	671	29.1	52.9	41.4	72.4
	Fixed	22.9	29.7	20.6	26.7	100.0	606	26.6	52.6	43.6	73.2
III	Comprehensive	20.8	29.2	20.8	29.2	100.0	1080	38.7	50.0	41.6	70.8
	De Facto	19.8	24.9	20.3	35.0	100.0	1053	44.1	44.7	40.1	65.0
	De Jure	21.1	29.2	20.8	28.9	100.0	1067	38.1	50.3	41.9	71.1
	Fixed	21.5	28.7	21.3	28.5	100.0	1040	37.1	50.2	42.8	71.5
IV	Comprehensive	29.9	35.5	15.8	18.7	100.0	784	22.1	65.2	45.7	81.0
	De Facto	30.8	31.7	18.4	19.0	100.0	640	19.9	62.5	49.2	80.9
	De Jure	29.8	35.4	15.8	18.6	100.0	780	22.2	65.1	45.6	81.2
	Fixed	29.9	35.3	15.7	18.7	100.0	799	23.3	65.2	45.6	80.9

Table 2

Percentage Distribution of Matching Results for the 1st and 2nd 12 months by Type of Enumeration - Births

Area	Method	Common Event C	Survey Only N_1	Registration Only N_2	Estimated by C-D Method E	$C + N_1 + N_2 + E = N$ N %	N (Abs.)	Standard Error of N	Total Coverage by Survey $C + N_1$	Total Coverage by Registration $C + N_2$	Total Coverage by both $C + N_1 + N_2$
2nd 12 Months											
I	Comprehensive	51.6	26.9	14.2	7.4	100.0	628	9.5	78.5	65.7	92.6
	De Facto	63.1	15.3	17.4	4.2	100.0	419	5.3	78.4	80.5	95.8
	De Jure	51.9	26.6	14.2	7.3	100.0	620	9.3	78.5	66.1	92.7
	Fixed	53.5	25.2	14.5	6.8	100.0	580	8.6	78.7	68.0	93.2
II	Comprehensive	40.1	32.4	15.2	12.3	100.0	624	13.8	72.5	55.3	87.7
	De Facto	39.9	30.4	16.8	12.8	100.0	463	12.2	70.3	56.7	87.1
	De Jure	40.4	32.0	15.4	12.2	100.0	616	13.7	72.4	55.8	87.8
	Fixed	40.6	31.7	15.5	12.2	100.0	611	13.5	72.3	56.1	87.8
III	Comprehensive	24.7	38.7	14.3	22.3	100.0	1312	34.4	63.4	39.0	77.7
	De Facto	26.0	37.2	15.1	21.6	100.0	1156	31.1	63.2	41.1	78.3
	De Jure	25.3	38.6	14.3	21.8	100.0	1267	33.1	63.9	39.6	78.2
	Fixed	25.4	38.5	14.3	21.7	100.0	1266	3.0	63.9	39.7	78.2

Table 3

Percentage Distribution of Matching Results for the 1st and 2nd 12 months by Type of Enumeration - Deaths

Area	Method	Common Event C	Survey Only N_1	Registration Only N_2	Estimated by C-D Method E	$C + N_1 + N_2 + E = N$ N %	N (Abs.)	Standard Error of N	Total Coverage by Survey $C + N_1$	Total Coverage by Registration $C + N_2$	Total Coverage by both $C + N_1 + N_2$
1st 12 months											
I	Comprehensive	33.1	14.8	36.0	16.1	100.0	175	9.2	47.9	69.1	83.9
	De Facto	36.3	12.6	37.8	13.2	100.0	127	6.9	48.9	74.1	86.7
	De Jure	33.5	14.4	36.4	15.7	100.0	173	9.0	47.9	69.9	84.3
	Fixed	33.5	14.4	36.4	15.7	100.0	173	9.0	47.9	69.9	84.3
II	Comprehensive	7.9	24.5	16.5	51.1	100.0	290	43.3	32.4	24.5	48.9
	De Facto	12.9	25.7	20.5	40.9	100.0	171	23.3	38.6	33.4	59.1
	De Jure	8.0	24.4	16.7	50.9	100.0	287	13.5	32.4	24.7	49.1
	Fixed	8.0	24.4	16.7	50.9	100.0	287	13.5	32.4	24.7	49.1
III	Comprehensive	7.8	23.1	17.5	51.6	100.0	486	56.7	30.9	25.3	48.4
	De Facto	9.1	23.9	18.5	48.4	100.0	372	44.5	33.0	27.6	51.5
	De Jure	8.0	23.5	17.4	51.2	100.0	477	55.2	31.5	25.4	48.9
	Fixed	8.0	23.5	17.4	51.2	100.0	477	55.2	31.5	25.4	48.9
IV	Comprehensive	22.2	34.4	17.0	26.3	100.0	229	16.5	56.6	39.2	73.6
	De Facto	26.8	28.2	21.9	23.0	100.0	142	11.1	55.0	48.7	76.9
	De Jure	22.7	35.2	16.5	25.5	100.0	224	15.9	57.9	39.2	74.4
	Fixed	22.7	35.2	16.5	25.5	100.0	224	15.9	57.9	39.2	74.4

Table 3

Percentage Distribution of Matching Results for the 1st and 2nd 12 months by Type of Enumeration - Deaths

Area	Method	Common Event C	Survey Only N_1	Registration Only N_2	Estimated by C-D Method E	$C + N_1 + N_2 + E = N$ N %	N (Abs.)	Standard Error of N	Total Coverage by Survey $C + N_1$	Total Coverage by Registration $C + N_2$	Total Coverage by both $C + N_1 + N_2$
2nd 12 months											
I	Comprehensive	32.9	26.9	22.1	18.1	100.0	185	10.1	59.8	55.0	81.9
	De Facto	38.4	19.2	28.3	14.2	100.0	110	6.4	57.6	66.7	85.9
	De Jure	32.9	26.9	22.1	18.1	100.0	186	10.1	59.8	55.0	81.9
	Fixed	32.9	26.9	22.1	18.1	100.0	186	10.1	59.8	55.0	81.9
II	Comprehensive	10.2	30.5	14.8	44.5	100.0	236	32.1	40.7	25.0	55.5
	De Facto	11.0	29.9	15.9	43.2	100.0	164	25.4	40.9	26.9	56.8
	De Jure	10.2	30.5	14.8	44.5	100.0	236	32.2	40.7	25.0	55.5
	Fixed	10.2	30.5	14.8	44.5	100.0	236	32.2	40.7	25.0	55.5
III	Comprehensive	10.5	37.3	11.5	40.7	100.0	418	40.3	47.8	22.0	59.3
	De Facto	12.4	35.1	13.7	38.8	100.0	307	31.0	47.5	26.1	61.2
	De Jure	10.6	37.3	11.5	40.6	100.0	416	39.9	47.9	22.1	59.4
	Fixed	10.6	37.3	11.5	40.6	100.0	416	39.9	47.9	22.1	59.4

Table 4A

Pattern of Rates of Vital Events Obtained by Survey and Registration Assistants in the First and Second 12 Months of Operation

Area	Period	Event	Ratio of $\frac{C + N_1}{C + N_2}$
I	1st 12 months	Births	0.89
		Deaths	0.69
	2nd 12 months	Births	1.19
		Deaths	1.09
II	1st 12 months	Births	1.42
		Deaths	1.32
	2nd 12 months	Births	1.31
		Deaths	1.63
III	1st 12 months	Births	1.20
		Deaths	1.22
	2nd 12 months	Births	1.63
		Deaths	2.17
IV	1st 12 months	Births	1.43
		Deaths	1.44

$C = N_1$ = Covered by Survey

$C = N_2$ = Covered by Registration Assistants

improved from 47.9% to 59.8% in Area I, the registration method slightly decreased from 69.1% to 55.0% and the overall combined method of survey and registration also slightly decreased from 83.9% in the first period to 81.9% in the second period. Slight improvements occurred during the second twelve months of operation over the first twelve months in Area II. The corresponding figures for Areas III and IV are also shown in Table 3. In general, the coverage rate for deaths is far below the coverage rate for births. This pattern has been noted in several surveys and sample registration systems in Africa where reporting of deaths has always been low.

6. Evaluation and Discussion of the Coverage Estimates

Table 4A shows the patterns of vital events obtained by survey and registration assistants in the first and second twelve months of operation in the four project areas. In Area I, during the first 12 months of operation, the registration assistants surpassed the survey enumeration team in netting births and deaths respectively. In all other areas the survey team had superior results, exceeding the registration team by ranges from 19% to 63% for births. These percentages increased in the second year of operation, possibly due to field supervision, improvements in the field techniques and knowledge of the area, although the same could be said of the vital registration assistants also. This finding is somewhat surprising due to the rather long six-month intervals between rounds during which memory lapse, migration, and other factors would have reduced the possibility of the survey enumerators obtaining most of the events. Assuming the figures of the survey team are correct, the conclusion to be drawn from the study thus far is that the survey team is more effective in collecting vital events

than the resident registration assistants. However, we are clearly uncertain about this conclusion. A more likely explanation is that there is over-reporting of children. Presumably the over-reporting is due to the enumeration of children under one year who were brought in from outside the project areas and who should be properly registered as new migrants into these areas.

The most illuminating comparisons are likely to be the cumulated first-birth rates compared with the proportion of women who are mothers, in order to detect under-reporting or over-reporting. The reporting of first births during a year will be affected by time scale errors (whether the event actually occurred within the period under study), but it is reasonable to expect that because of the youth of most mothers and the importance of a first birth, omissions and age differentials in the reference period will have little influence. The classification of a woman as a mother will be less subject to error than the determination of the number of children born to her. For these reasons and because the proportion of women who become mothers after the age of 30 years is small in most African populations, the ratio of ever-fertile women to cumulated first birth rates at ages 25-29 should be a useful index of the time scale error. The data on proportions who become mothers in the population was obtained from the independent fertility survey in 1971 and the first order birth rates are obtained from vital event registration of current births. Table 4B shows this comparison by areas for the project area.

The evidence from the P/F ratios (not shown here) for the younger women suggests that in the four areas, 2 per cent to 20 per cent over-reporting of births occurred in the current year. These types of discrepancies were further checked by other evaluation procedures. The age

Table 4B

Proportion of Women Who are Mothers: Observed (0)

and Estimated from First Order Births (S)

Area	Age Group at	Proportion of Women who are mothers (0)	(S) 1*	(S) 2*	Ratio 0/S 1*	Ratio 0/S 2*	Chandrasekhar-Deming Correction to Births 1*	Chandrasekhar-Deming Correction to Births 2*
I	25-29	0.992	0.8210	0.6455	1.21	1.54	1.23	1.08
	20-24	0.839			1.02	1.30		
II	25-29	0.959	0.8715	0.9185	1.10	1.04	1.36	1.15
	20-24	0.869			1.00	0.95		
III	25-29	0.973	1.2490	1.0780	0.77	0.90	1.41	1.29
	20-24	0.858			0.69	0.79		
IV	25-29	0.993	1.0050	-	0.99		1.23	
	20-24	0.854			0.85			

*1 - First twelve months

2 - Second twelve months

specific first birth rate was calculated to give a measure analogous to the total fertility for all births; the resulting index would be a current estimate of 0, the proportion of women who become mothers. The table shows the estimate of 0 found for the areas. As Areas III and IV give measures for the proportion of women becoming mothers well over 1.00, particularly for Area III, the first births reported in the current year are clearly too high. An indication of the corrections needed for over-reporting can be obtained by dividing the proportion of mothers in the age group of women 20-24 or 25-29 years by the corresponding estimation of S shown in Table 4B.

There is some distortion from age errors in the earlier part of the reproductive period but the calculations for first order births give a clear and

simple demonstration of the method of analyzing the cumulative first order birth rates and the proportion of women who become mothers. In the interpretation, it is necessary to allow for the fact that the numbers of first births are those recorded by both systems but that the adjustment for the events on either list is not included. It seems the proportions who become mothers as derived from the survey by age 25-29 are far too high (suggesting over-reporting), and the proportions of women who become mothers derived from first order birth rates are far too low, particularly for Area I in the second twelve months of operation. In Area III first order birth rates are simply too high. For Area I, the O/S ratios imply that the number of first order births would have to be raised by some 21 per cent in the first 12 months to give agreement with the directly recorded proportions. For Area III, the first births would have to be decreased by something of the order of 23 per cent for reasonable agreement to be achieved.

There is some agreement for Area I with the Chandrasekhar adjustment which gives some confidence in the results, whereas in other cases the Chandrasekhar adjustment is either too far short of closing the gaps or too wide. The variations by area are mainly notable for better coverage for Areas I, II and IV and poorer coverage in Area III in the first 12 months of operation. When the proportion of women who become mothers by age 20-24 was used, the results did not differ significantly. It might have been expected that a high O/S ratio would be associated with a high Chandrasekhar correction factor (since poor performance in the component collection procedures gives greater scope for the bias due to non-independence) but there is no evidence of this, on the other hand, the reverse seems to occur. This is borne out by the low ratios in Area III for O/S values and the high ratios for the Chandrasekhar correction factor.

The validity of the above estimates will depend however, on the ful-

fillment of the following primary conditions:

a) the matching procedure successfully identifies all true matches;

b) all events identified in either of the two systems are true events, (i.e. occurred in the population under investigation and in the appropriate time period);

c) the two systems are independent, (i.e. the probability of an event being omitted from one system is not related to the chance of that event being omitted from the other system).

Deviations from the above conditions may seriously affect the accuracy of the estimates derived from the method. Yet these conditions are quite stringent and the degree of deviation from their fulfillment is at best difficult to ascertain. First of all, the notion of what constitutes a proper match is ambiguous in almost all practical situations. In general, the more stringent the definition for a match, (i.e. the larger the number of attributes that coincide in order to establish a 'true' match) the smaller will be the estimate of C and the larger will be N_1 (events covered by survey) and N_2 (events covered by registration), consequently the larger will be the estimate of N (total estimate of events).

There is the danger that overly stringent matching criteria will result in inflated estimates of the true number of events, apart from making the matching process especially laborious. On the other hand, an estimate based on loose matching criteria may yield an underestimate. The task of finding the mean between these two extremes is difficult. The obvious solution, that of investigating every suspected match in detail, is limited by cost considerations. Usually simple rules will have to be imposed but the fact that such rules must necessarily take account of peculiarities of the specific situation makes any generalization about them difficult.

Most reports of the scientific experience from other projects have been cautious on these points although the same cannot be said about the presentation

of the substantive results. There is, however, much scope for genuine disagreement about the importance of particular defects and the likelihood that they may be overcome by improved field or analytic techniques. Jabine and Bershad (1970) have given theoretical consideration to the effects of biases and sampling errors on the independent matching period. They point out that the effects of non-independence on the estimates may be large, particularly when the proportions of events recorded are not high. On the other hand, Brass, W. (1969) has pointed out in a review of the C-D methods that improved performance from the C-D method (shown sometimes by the small proportion of events missed by both methods) in some cases may be due to duplication and not to independent matching criteria. Brass further suggested that the latter may not be a crucial feature in a matching experiment.

A useful review of Asian experience covering Pakistan, Turkey, Korea, Thailand and several surveys in India is given by Selzer, W. (1969). The summary of Selzer's examination of nineteen studies is presented in Table 4C.

Table 4C

Median and Range of Estimated Percentage of Events Covered, Selected Asian PGE Studies

Type of System	Number of Systems	Births		Deaths	
		Median	Range	Median	Range
All	38	77	28-96	69	23-91
Civil Registration	6	56	43-86	50	33-92
Special Registration	12	73	62-92	66	50-91
One-Visit Survey	7	67	28-96	51	23-90
Serial Visit Survey	13	83	66-92	72	50-89

Source Selzer, W. (1969) Some Results from Asian Population Growth Studies, Population Studies 23, Table 3, p. 403.

One striking conclusion one can obtain from this survey of other experiences is the failure of any one of the systems to achieve anywhere near complete coverage with consistency. The second important conclusion is that birthsare somewhatbetter recorded than deaths. Independence is a necessary, but by no means sufficient, condition for success of collecting vital events by the C-D Method. The conclusions obtained in this project, therefore, seem to reinforce earlier conclusions arrived at in other parts of the world on the application of the C-D for the estimation of vital rates.

7. Analysis of Data on Pregnancy Records

A useful device which has been employed in ensuring the accuracy of the data on live births and infant deaths is an inclusion of a question regarding current pregnancies*. This information is compiled by the registration assistants who follow up these pregnant women on a continuing basis, and by the survey team in the succeeding rounds who record the outcome of the pregnancies. A major problem which arises in the follow-up of these pregnant women is the loss of women through migration and probably death. These problems are discussed in the light of the coverage estimates. The results of the analysis are presented in Tables 5 and 6. Comparative data is also presented in Table 7 from an independent fertility sample survey carried out in 1971 in all of the project areas.

---

*The question on current pregnancies was once rejected in Ghana and deleted from a national sample survey by the Central Bureau of Statistics, Ghana, because it was believed that in Ghana such a question may be offensive. Our experience has shown that this is not so. See U.N. Statistical Papers M No. 51 - Demography of Sample Surveys, page 74 (1971).

Table 6A

Percentage Distribution of the Summary on Pregnancy

Outcome in Round II by Areas

Area	Outcome of Pregnancies							
	Born Alive and now Living	Born Alive and now Deceased	Moved Away After Birth	Still Birth	Miscarriage	Outcome Unknown	Total Pregnancies Recorded Percent	Total Pregnancies Recorded Absolute N
I	76.8	2.6	4.8	1.3	3.7	10.8	100.0	379
II	73.5	3.8	9.5	1.7	3.4	8.1	100.0	234
III	60.7	3.4	7.1	2.0	5.1	21.7	100.0	351
IV	58.2	3.0	7.3	1.7	3.4	26.3	100.0	232
Totals	67.8	3.2	6.9	1.7	4.0	16.4	100.0	1,196

Notes: During the Round 3 recensus (March-July 1973) some fertile females reported additional children and miscarriages. The pregnancies which resulted in these births and miscarriages had not been previously reported (i.e. they did not include pregnancies reported during Round 2). The women were either entering the project villages for the first time with the pregnancies or did not report the pregnancies to the enumeration team during Round 2. A few of the cases might have been missed by the enumerators. The outcome of these pregnancies are displayed in Table 6B as outcome of pregnancies not recorded during Round 2 (source: Round 3).

Table 5

Pregnancies Recorded by Survey

Area	Round 2	Round 3	Total
I	379	142	521
II	234	225	459
III	351	383	734
IV	233	319	552
Total	1197	1069	2266

Note: Round 2 refers to re-census of March-July 1972, Round 3 refers to another recensus of April-July 1973 in the project area.

Table 7

Percentage of Women Recorded Pregnant from

Vital Event Surveys, Recensuses and Fertility Surveys

Source		Area			
		I	II	III	IV
a) 1971 Fertility Survey		10.7	12.4	18.4	11.6
b) Rounds 2 and 3*		15.6	16.5	18.6	14.5
c) Percentage of Women in the 15-49 age group delivering live birth	1+	19.9	21.9	24.3	20.9
	2+	17.7	20.3	24.6	-
d) Difference between b) and c)		4.3	3.8	6.0	6.4

*Women interviewed in the fertility survey were age 15-44 whereas women questioned in Rounds 2 and 3 were age 15-49.

1+ First 12 months

2+ Second 12 months

The essential features of Table 6A are clear. Note in Table 6A the high percentages of the results of the pregnancy record for outcomes unknown: 21.7 per cent for Area III and 26.3 for Area IV. This may partially explain the poor outcome of the results of matching in Areas III and IV as most of these events whose outcomes were still unknown could have matched for the common event between the two systems. This also explains the comparatively low figures in the born alive and now living group for Areas III and IV in Table 6. There is no doubt that migration is a factor in these low figures, as may be deaths, whose coverage are equally low for Areas III and IV. The recorded pregnancies in both rounds 2 and 3 are higher for Areas III and IV, which is possibly an indication of high incidence of births in these two areas. However, the more frequent the events of birth and death, the more likely is the chance that

Table 6B

Percentage Distribution of the Outcome of Pregnancies
Not on Pregnancy Diary for Round 2 by Areas

Area	Outcome of Pregnancies							
	Born Alive and now Living	Born Alive and now Deceased	Moved Away After Birth	Still Birth	Miscarriage	Outcome Unknown	Total Pregnancies Recorded Percent	Total Pregnancies Recorded Absolute N
I	87.1	5.4	2.9	1.3	2.5	0.8	100.0	241
II	88.7	2.6	1.1	2.8	4.8	0.0	100.0	470
III	87.4	3.8	0.2	1.7	6.9	0.0	100.0	836
IV	87.0	4.1	1.3	2.2	5.3	0.1	100.0	536
Total	87.5	3.8	1.0	2.0	5.6	0.1	100.0	2,182

Source Round 3