PN-AAP-709
ISN 34178

ECONOMIC ANALYSIS OF AGRICULTURAL POLICIES

RSSA 3-77

BOLIVIAN RURAL HOUSEHOLD SURVEY

BY GARRIE J. LOSEE,
Deputy Associate Director,
National Center for Health Statistics, DHEW
when serving as Sampling Advisor to USAID/
Bolivia under the sponsorship of the Office of Nutrition Economics,
OICD, USDA.

BOLIVIAN
RURAL HOUSEHOLD SURVEY

PROPOSED SAMPLE DESIGN *

The sample design proposed for the Bolivian Rural Household Survey (RHS) is a highly stratified two-stage probability sample of clusters of rural households in Bolivia. The design has two levels of sample size: Level I which assumes a total survey budget of $550,000; and Level II which assumes a total survey budget of $350,000. The sample size of Level I is 3,120 rural households and of Level II is 1,440 rural households.

Target Population--The population to be surveyed is the population living in areas defined as being rural areas during the 1976 Population Census. These areas had a population of less than 2,000 during the 1976 Population Census.

Primary Sampling Units--The 93 Provincias which constitute the geographic area of Bolivia are proposed as the primary sampling units (PSU's). The rural population of the Provincias ranges from 5,000 to 70,000 persons but most are in the range between 20,000 and 40,000 persons. Both the Level I and Level II designs call for a highly stratified selection of one PSU from each of 30 strata. The 30 strata will be formed by classifying the 93 Provincias into strata according to their Department, ecological zone, and major crop production. The selection of the sample PSU's will follow a Controlled Selection Technique in which the selection will be controlled to population totals for population density and population growth between two recent Censuses.

Second Stage Units--The second stage units are compact clusters of an expected four rural households. A sample cluster is selected by a process which includes first selecting a Census enumeration district (or segment). Each segment includes on the average about 40 households in rural areas. The Level I design calls for the selection of 26 segments from within each sample PSU and then a cluster of an expected four households from each segment. The Level II design calls for the selection of 12 segments instead of 26.

Listing of Households--All households in selected segments are to be listed independently of interviewing. Compact clusters of four households will be constructed in the central office and one will be randomly selected to be a sample cluster.

Discussion of the Proposed Design

The proposed design is a considerable departure from that of the Bolivian Socio-Economic Farm Survey. Since the proposed design will be compared to the Farm Survey design, it is well to review the reasons for proposing the

---

* Proposed by Garrie J. Losee, Deputy Associate Director, National Center for Health Statistics, DHEW, when serving as Sampling Advisor to USAID/Bolivia under the sponsorship of the Office of Nutrition Economics, OICD, USDA.

Best Available Document

the above design, and then citing its advantages and disadvantages as compared to a design with segments as the PSU's and mobile teams of interviewers.

The first approach considered in reaching a reasonable design for the RFS was identical to that of the Farm Survey. A total variable budget of $300,000 was assumed to be available. The cost equation for the seven-day RFS was

$$C = C_1 \sqrt{Am} + C_2 m + C_3 m\bar{n}$$

$$300{,}000 = 1{,}237 \sqrt{m} + 85m + 71m\bar{n}$$

An explanation of the unit costs is appended to this report. The values of m (segments) and $\bar{n}$ (households) that fit the above equation and the precision they produce in terms of relative variance are shown in the table below for an estimated rate of 0.50 and a 8=0.353:

$\bar{n}$	m	$m\bar{n}$	$V_Y^2 = \frac{\hat{V}^2}{m\bar{n}} [1 + 8 (\bar{n} - 1)]$
1	1600	1600	0.00062
2	1200	2400	0.00056
3	900	2700	0.00063
4	800	3200	0.00064
6	560	3360	0.00082
8	450	3600	0.00096

The optium design therefore is a design of $\bar{n}$=2 households selected from m=1200 segments PSU's. However, even an $\bar{n}$ of 4 which is close to optium is not a feasible PSU workload for a team of four interviewers in which an interviewer is supposed to complete 2 households simultaneously during a 7-day period.

This lead to the hypothesis that a reasonable design would be based on a pair of interviewers completing a number of clusters of four households over a one-year period in a large sample PSU where travel distances from home to each sample cluster would not be too large. The Provincia seemed to be the most reasonable geopolitical unit that fit those criteria. It should be noted that it is planned that the interviewer will be women and that their working in pairs is desirable as a safety precaution. Also, the interviewers would only be paid on the days they were working.

Since a secondary but important aspect of the RIS is to measure seasonal variations in food intake and nutritional status, an equal distribution of the sample over the weeks of the year was considered. A design which allocated a cluster of four sample households to every other week seemed to have considerable merit. The unit costs of such a design were estimated and are appended to the

report. The cost equation is:

$$C = C_1 m\bar{n}\sqrt{A} + C_2 m + C_3 m\bar{n} + C_4 m\bar{n}\bar{\bar{q}}$$

$$C = 1.46\ (30)(26)\sqrt{1600} + 570\ (30) + 74.6\ (30)(26) + 50\ (30)(26)(4)$$

$$C = 45{,}652 + 17{,}100 + 58{,}188 + 156{,}000$$

$$= 276{,}940$$

The precision of the design for an $\gamma = 0.50$ is estimated by the equation:

$$V_\gamma^2 = \frac{\hat{v}^2}{m\bar{n}\bar{\bar{q}}}\ [\delta_1\ \bar{n}\bar{\bar{q}} + 1 + \delta_2\ (\bar{\bar{q}} - 1)]$$

$$V_\gamma^2 = \frac{1}{3120}\ [0.03\ (104) + 1 + 0.300\ (3)]$$

$$= 0.0016$$

and $V\gamma = 0.04$ or 4.0 percent. If the estimate of γ is 0.50, the chances are 19 out of 20 that the population value of γ is greater than 0.46 and less than 0.54.

The next step was to compare the 30-PSU sample design described above to a Farm Survey type design having a total variable cost of $276,940 and a mobile team with only two instead of four interviewers and four households being completed in a 7-day period. The cost equation is:

$$C = C_1\sqrt{mA} + C_2 m + C_3 m\bar{n}$$

$$C = 1{,}237\sqrt{m} + 85m + 97m(4)$$

$$276{,}940 = 1{,}237\sqrt{m} + 473m$$

Solving, $m = 525$ and $n = m\bar{n} = 2100$ households.

The precision of the above design for an $\gamma = 0.50$ is estimated to be:

$$V_\gamma^2 = \frac{\hat{V}}{m\bar{n}} [1 + 8 (\bar{n} - 1)]$$

$$V_\gamma^2 = \frac{1}{2100} [1 + 0.353(3)]$$

$$= 0.00098$$

and $V\gamma = 0.031$, slightly lower than the 30-PSU design.

Another comparison was made to a Farm Survey type design with $\bar{n}=8$ households. In that case the cost equation is:

$$276{,}940 = 1{,}237\sqrt{m} = 85m + 71m(8)$$

Solving, $m = 385$ and $n = m\bar{n} = 3080$ households.

Its precision is estimated to be:

$$V_\gamma^2 = \frac{1}{3080} [1 + 0.353(7)] = 0.00113$$

and $V\gamma = 0.034$, again slightly lower than the 30-PSU design.

The final comparison was made as to the accuracy of the three designs in estimating seasonal characteristics. The results were:

Design	n	V_γ^2	$V\gamma$
30-PSU	780	0.00344	0.059
Farm Survey, $\bar{n} = 8$	770	0.00452	0.068
Farm Survey, $\bar{n} = 4$	525	0.00372	0.062

Here, the 30-PSU has a slight advantage over the other two designs.

In concluding this discussion, the advantages and disadvantages of the 30-PSU design are:

Advantages:

- Interviewers can be recruited from within their work area (Provincia).
- It can be used as the rural part of a national household for a variety of surveys.
- Slightly higher precision for seasonal estimates.
- Not highly dependent on vehicles.
- Listing of household done separately from interviewing.
- Flexibility exists to extend the 7-day assignment period to 14 days if there are more than the expected four households in the sample cluster.

Disadvantages:

- Supervision is not as direct as in other two designs.
- Slightly lower precision for annual estimates (unless there is large seasonal variation in parameters being measured).
- Logistics and communications are complex.

Level II Sample Design--The sample design proposed for Level II is also a 30-PSU design, but due to the limited budget, it would have a sample of only 1,440 households. The reduction from Level I would be in the number of clusters of 4 sample households in each sample PSU. The Level II sample would have 12 such clusters instead of the 26 in Level I. These would be distributed one to each of the 12 months. The total variable cost for the Level II design is estimated to be $137,000.

The precision of the proposed Level II design is estimated to be such that an estimate of a rate of 0.50 based on the 1-year sample would have a relative sampling error of 4.8 percent. A comparable estimate for a 3-month sample would have a relative sampling error of 9.3 percent, an error which would make it unlikely to measure small or even moderate seasonal changes in field intake or nutritional status.

ALTERNATIVE CONSIDERATION

The proposed sample design and total sample size are largely dependent on the method planned for obtaining accurate measurements of food intake. The method requires the weighing of food twice a day over a 7-day period. Although this method is the preferred method from the standpoint of accuracy of measurements, it is both costly and difficult to carry out effectively in a large-scale sample survey. Accordingly a Pilot Study will be undertaken to consider the advantages and disadvantages of a similar method based on 3-day measurements instead of seven. The possible consequences of a decision to opt for a 3-day method are accordingly examined below.

1. Retaining same sample design and size of sample

   The main consequence of this decision would be to reduce to unit cost of a sample household from $50 to $30, and the total variable cost from $276,940 to $214,540 (Level I) or $137,000 to $109,000 (Level II). The sampling reliability would remain the same with the overall.

2. Keeping budget constant and increasing sample size

   The obvious consequence of this decision is to improve the sampling reliability of both annual and seasonal estimates. The increase in sample size would be in terms of clusters of 4 sample households. The increase for Level I would be from 26 to 43 clusters per PSU or from 3,120 sample households to 5,160. The increase for Level II would be from 12 to 20 clusters per PSU or from 1,440 sample households to 2,400. The resulting effect on sampling reliability would be, in the case of Level I, to reduce the relative sampling error by about 10 percent (from 4.0 percent to 3.6 percent for an estimate of r = 0.50). The effect on Level II is similar; the reduction in relative sampling error being slightly over 10 percent (from 4.8 percent to 4.2 percent for a rate of 0.50).

   The ability to make seasonal comparisons would be considerably improved. The reduction in relative sampling error for Level I would be from 5.9 percent to 4.9 percent and for Level II from 9.3 percent to 6.5 percent; both again for an estimated rate of 0.50.

THE PILOT STUDY

The Pilot Study is the most critical element of the Rural Household Survey. Normally, pilot surveys are helpful in refining questionnaires and procedures, as this Pilot Study will. However, due to the innovativeness of the methods of information gathering built into the Rural Household Survey and the demand they put on the field apparatus, the Pilot Study should also be the final decision point for going ahead with the Rural Household Survey or cancelling it. This decision should be based on the culmination of evidence from the Pilot Study regarding:

1. the level of nonresponse, particularly refusals;

2. the ability of the field staff to carry out their assigned responsibilities;

3. the completeness of the data; and

4. the face validity of the survey design and the information gathering methods.

On a more positive note it is assumed that, although problems will be encountered they will be resolvable and the decision will be to go ahead with the Rural Household Survey. In that case, the purposes of the Pilot Study are to:

1. determine the level and sources of nonresponse and developed methods to decrease nonresponse in the RHS;

2. learn of any problems remaining in the questionnaire, either for the interviewer or the respondent, and correct them before the RHS;

3. see how well the field procedures work and which ones should be improved and how;

4. determine the respondent burden and how it can be reduced;

5. obtain some measures of intraclass correlations;

6. evaluate the efficiency of alternative methods of measuring food intake; and

7. select the best interviewers as field supervisors for the RHS.

The Pilot Study will consist of 100 households selected in 26 clusters of 4 households. Thirteen pairs of interviewers will each be assigned two clusters of households. The Pilot Study interviewing will take place over a 3-week period, preceded by a one-week training session. INE staff will supervise and observe the field activities. The interviewers will be located at a hotel in a central urban center or at their home, if recruited locally (as some should be). Every attempt should be made to simulate the RHS within the limitations of time and a single Provincia.

RURAL HOUSEHOLD SURVEY PROJECT

ALTERNATIVE 3 - SCALED-DOWN SURVEY

A. Purpose--To achieve bare minimum information goals set by USAID/Bolivia for the project deadline (August 31, 1981) using basic proven methodology, and building upon the experience and materials of the Bolivian Socio-Economic Farm Survey.

B. Time Schedule--The two-year project would begin September 1, 1979 and end August 31, 1981. The major activities are:

1.	Questionnaire and field procedures development (6 months)	9/1/79 - 3/1/80
2.	Questionnaire and procedures pretest (50 households; 3 teams)	3/3/79 - 3/21/80
3.	Revisions to questionnaire and field procedures	3/22/80 - 5/24/80
4.	Training session (5 days)	5/27/80 - 5/31/80
5.	Survey Period I (8½ weeks)	6/3/80 - 8/1/80
6.	Survey Period II (8½ weeks)	9/30/80 - 10/28/80
7.	Survey Period III (8½ weeks)	2/3/81 - 3/3/81
8.	Data processing	8/2/80 - 5/30/81
9.	Tabulation/analysis	6/2/81 - 8/31/81

C. Duration of Interview--average of 4 hours in one day for each household.

D. Size of Sample--

1. n - 2,160 households, selected in clusters of 8 households from within 270 PSU's (Census Enumeration Districts). About 200 of these PSU's are expected to be the same PSU's already listed for the Farm Survey. Thirty will be selected from the Departments of Beni and Pando.

2. fieldwork to be conducted by 6 mobile teams each consisting of a driver, a supervisor, and 4 interviewers. Including travel time between PSU's and a day of rest every 8 days, it is necessary to allocate 4 days of a team for each segment of 8 households.

3. workload/team = 8 households x 15 segments x 3 periods.

E. Administration--Project to be directed by MACA's Office of Sector Planning. Sample selection, fieldwork, coding and editing, and possibly data entry, mechanical editing and tabulation to be done also by MACA. Data entry, mechanical editing, and tabulation possibly to be done by subcontract with INE.

F. Cost Element Estimates--

1. Data Collection

• Interviewer and Supervision costs Team cost per day x days per segment x segments ($70 x 4 x 270)	$74,600
• Mileage costs (18,000 miles at $.30/mile)	5,400
• Segment listing ($80 per segment x 70 segments)	5,600
• Pretest ($54 per household x 50 households)	2,700
• Training session ($55 per team/day x 6 days)	2,000
Total	90,300
2. Data Processing ($20 x 2,160 households)	43,200
3. Tabulations	20,000
4. TDY's	
• BUCEN (6 months - 10 trips)	40,000
• Analyst (9 months - 4 trips)	60,000
Total	100,000
5. Communications, Miscellaneous	6,500
6. Field Director (MACA) - 18 months	50,000
Total Estimated Cost	$310,000

10

RURAL HOUSEHOLD SURVEY PROJECT

ALTERNATIVE 2 - PILOT SURVEY

A. Purpose--To develop and evaluate alternative survey methodologies for measuring household expenditures and food consumption in the rural zone of Bolivia. The product would be an optimal method, together with questionnaires and field procedures, for undertaking a large-scale rural household survey method.

B. Time Schedule--The one-year project would begin September 1, 1979 and end by August 31, 1980. The major activities are:

1.	Development of data collection methods and instruments (6 months)	9/1/79 - 3/1/80
2.	Training session (7 days)	3/3/80 - 3/9/80
3.	Survey period (3 weeks)	3/11/80 - 3/30/80
4.	Data processing/analysis (5 months)	4/1/80 - 8/31/80

C. Duration of Interview--Average of 2 hours per day over a 7-day period for each household.

D. Size of Pilot Survey--

1. n = 200 households, selected in clusters of 4 households from within 12 segments (Census Enumeration Districts) in each of 4 Provinces.

2. fieldwork to be conducted by 48 interviewers; 12 working in pairs in each of 4 Provinces and to be supervised by 8 supervisors.

3. workload/interviewer = 2 households x 2 segments
   = 4 households

E. Administration--Project to be directed by MACA's Office of Sector Planning, with fieldwork and data processing contracted with INE.

11

F. Cost Elements--

1. Data Collection

• Interviewers (48 x 20 days x $10/day)	$9,600
• Segment listing (48 segments)	2,250
• Supervisors (8 x 20 days x $15/day)	2,400
• Drivers (8 x 20 days x $15/day)	2,400
• Jeeps, transportation costs	2,500
• Training session (48 x 10 days x $10/day) + (8 x 10 days x $15/day)	6,000
Total	25,150

2. Data Processing (200 x 20)	4,000
3. Tabulations	10,000
4. TDY's	
• BUCEN (6 months - 10 trips)	42,000
• Analyst (3 months - 4 trips)	20,000
Total	62,000
5. Local Salaries	
• MACA - 2 positions/12 months	40,000
• INE - 4 positions/12 months	60,000
Total	100,000
Total Cost	$201,150

G. Balance of Funds--The residual budget allocation of $150,000 allocated to this project would be reprogrammed to accelerate the analysis of data from the companion agricultural surveys to this project.

C_4 = COST PER HOUSEHOLD (7-day method)

Interviewer time - $10 per day; 8 days per 2 households	$40.00
Manual editing and coding	2.00
Data entry, editing, and tabulations	7.30
Printing of questionnaires, forms	.70
	50.00

C_3 = COST PER SEGMENT

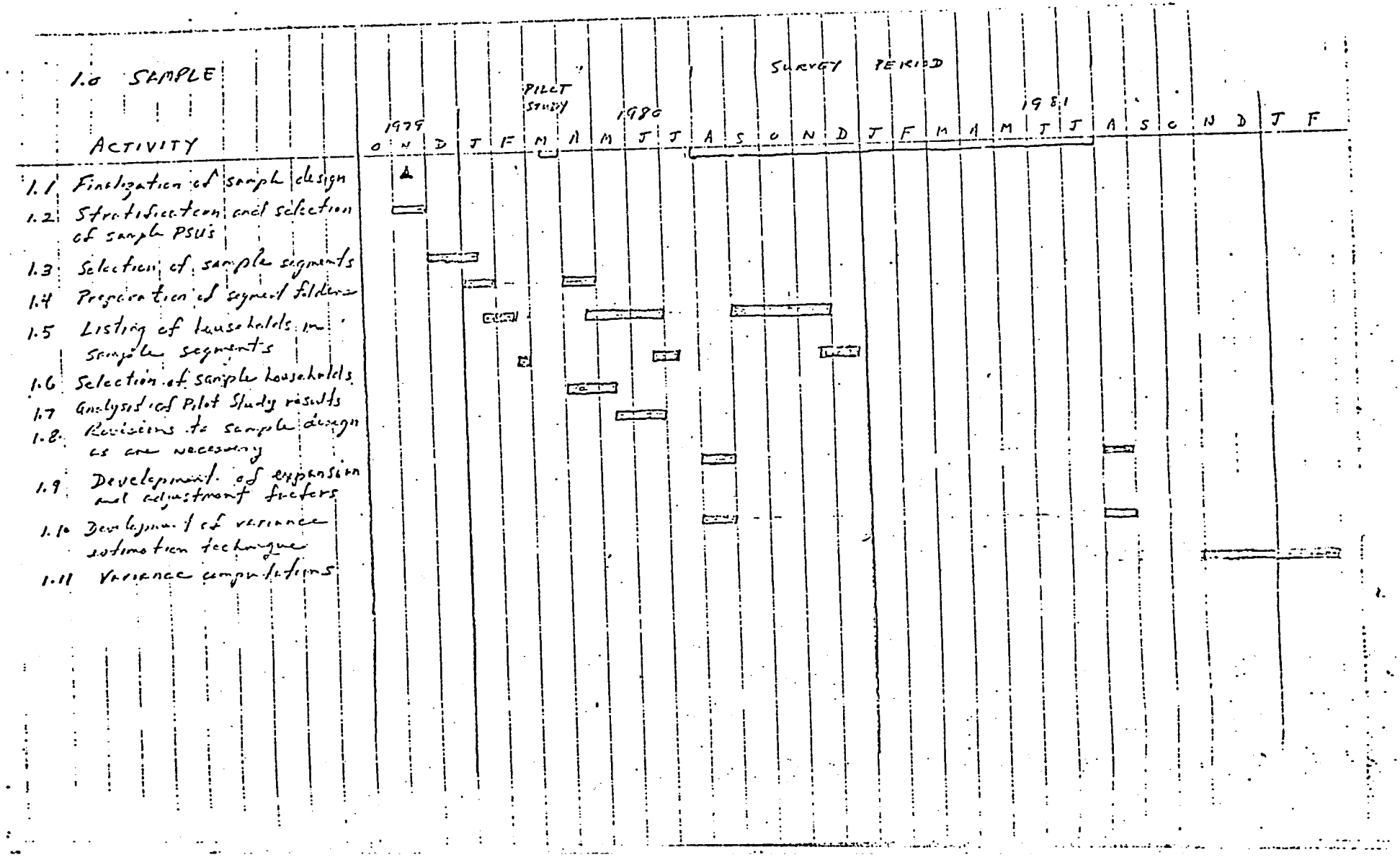
Mapping and segment folder		10.00
Listing of households (30 days for 13 segments at $15/day)		34.60
Supervision		
Driver	1 day at $15/day	15.00
Supervisor	1 day at $15/day	15.00
		74.60

C_2 = COST PER PSU

Sample selection		20.00
Communications		100.00
Driver	15 days at $15/day	225.00
Supervisor	15 days at $15/day	225.00
		570.00

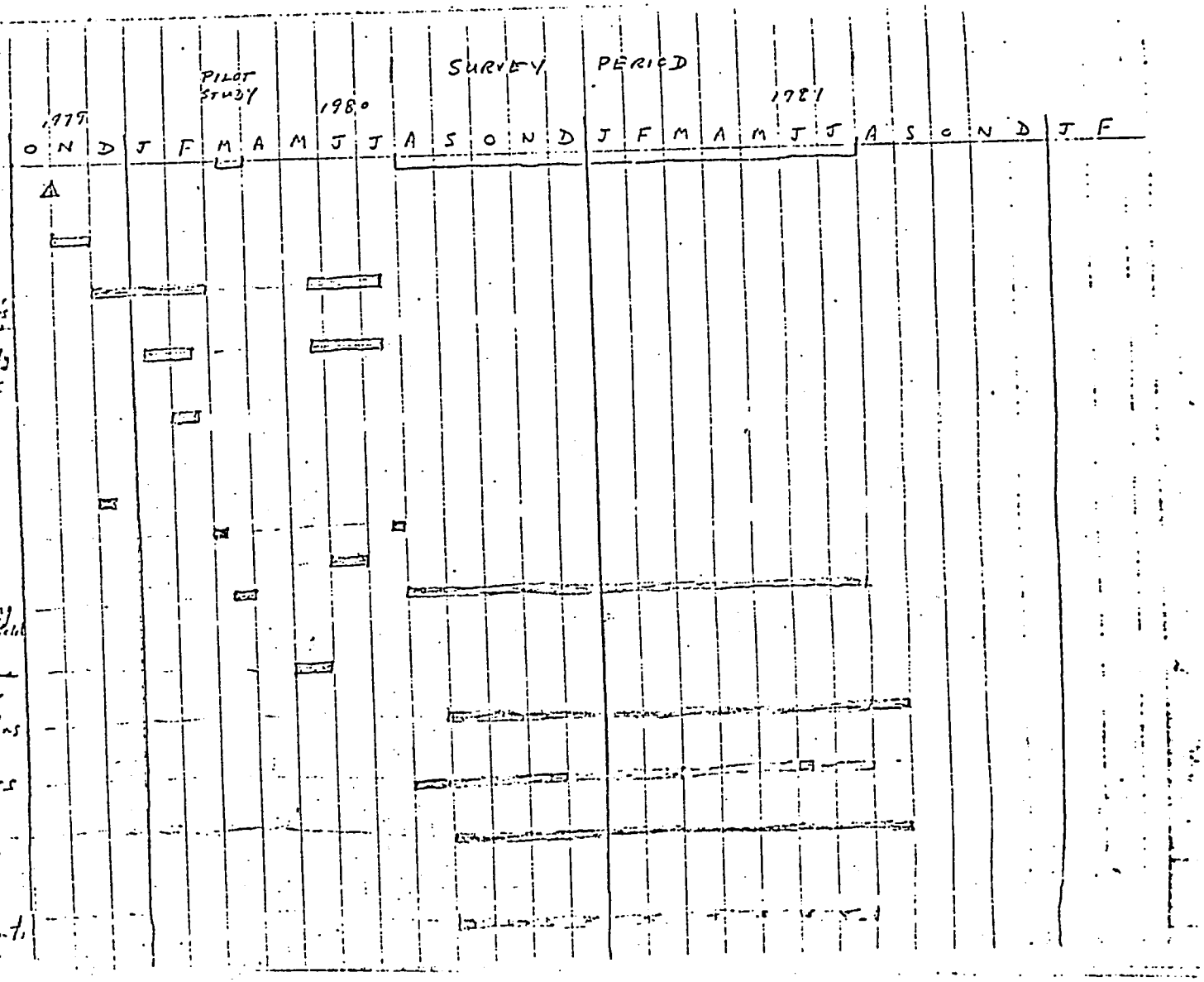
C_1 = SUPERVISORY TRAVEL (per mile)

Jeep mileage	0.30
Driver	0.58
Supervisor	0.58
	1.46

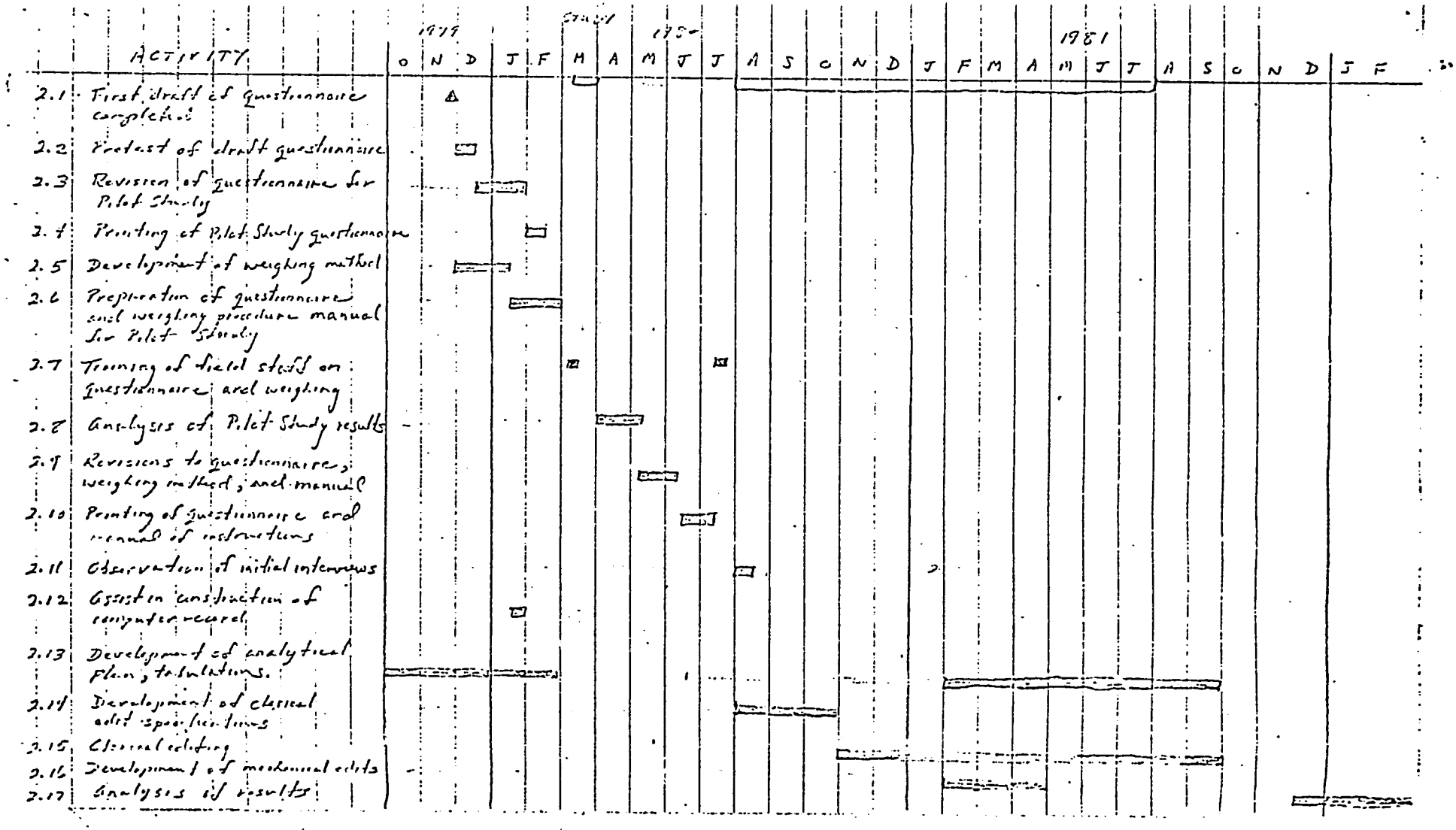


Best Available Document

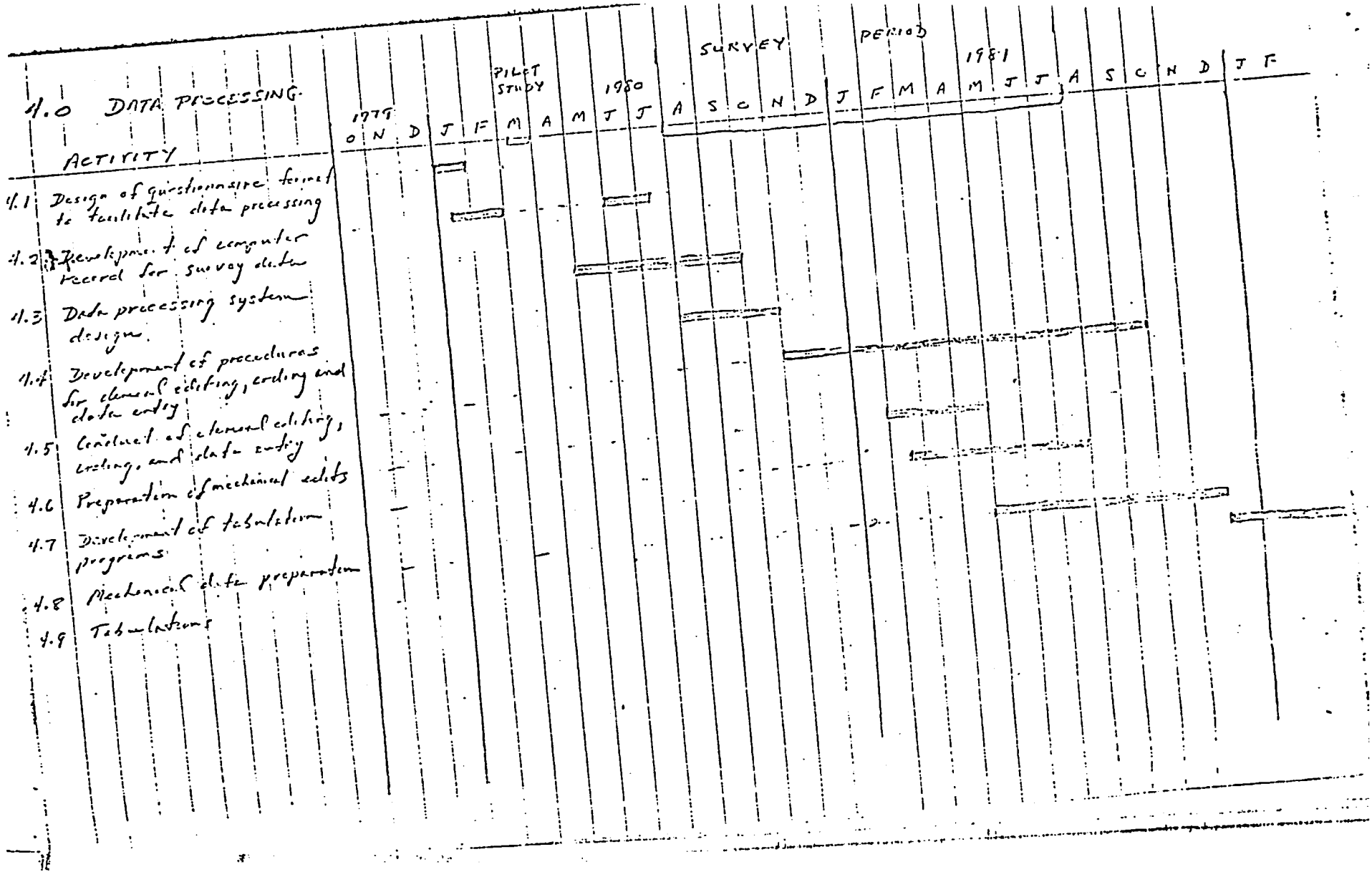
3.0 Field Operations



Best Available Document



Best Available Document



Best Available Document