EVALUATING ALTERNATIVE METHODS FOR DETERMINING OVERLAP DOMAIN IN MULTIPLE FRAME SURVEYS*

Raymond R. Bosecker, USDA

Introduction

Determining reporting units which are common to different sampling frames is not a simple task. To avoid omission or duplication when combining data from separate frames, the survey design must contain clear and concise rules of association. The existence of complex sampling units comprised of two or more individual units necessitates special rules for which there are various alternatives. This study addresses the concepts, problems and advantages associated with three alternative procedures which may be used to define the overlap domain between frames.

USDA livestock estimates are based on data from two sampling frames -- area and list. The area frame is complete for the population of interest. The list frame, while incomplete, may be stratified according to a measure of size for the variable being surveyed. The sampling units on the list are of three types: 1) names of individuals, 2) combinations of individuals (partnerships), and 3) names of farm operations. This is a common situation for lists of names which are to represent businesses. The problem for a multiple frame application arises from associating the reporting units, also composed of three types, back to the sampling units and matching between frames. Partnerships and operation names complicate the matching process because the same individual may be involved in operations of these types as well as his own individual business. Three alternatives investigated to determine overlap between frames given the nature of the list were: I) Partial Nonoverlap Procedure, II) Maximum Overlap Procedure, and III) Exact Match Procedure.

Partial Nonoverlap Procedure (Alternative I)

The primary purpose of this procedure was to minimize the effects of partnership operations on the sampling errors. This procedure relied on some basic assumptions:

a. Each partner in a partnership will report for the entire operation whether contacted through the area or list sampling frames.
b. Each partner will also report his individual operation if there is one.
c. Each partner will correctly identify all of the other partners.
d. Each partner whose name appears in the list frame will be identified.

Figure I is a decision diagram used to prorate partnership operations using the partial nonoverlap procedure. If an area frame tract is a partnership operation, the partnership has a name and the name is on the list, the tract is overlap. If a partnership operation does not have a name, the procedure is to determine if any or all of the partnership or corporate members are on the list. If all members are on the list, the tract is overlap. If only a fraction of the total partners are on the list, a partial nonoverlap fraction is computed. The nonoverlap fraction is one minus the number of partners on the list divided by the total number of partners. This means it is necessary to rely on the assumption that all people listed as partners are in fact partners by the survey rules and definitions.

The procedure for editing data reported by a partnership operation selected from the list frame must be compatible with that used to derive a nonoverlap fraction. For example, consider the case where an individual name is selected from the list and the respondent indicates a partnership arrangement on the questionnaire. If a partnership name is reported by the individual and it is also on the list, then all partnership data are to be removed from the individual's questionnaire since the partnership operation had an independent chance of being selected. If there is no partnership or corporate name, a portion of the partnership data is included on the individual's questionnaire. Since there is more than one chance of selecting the partnership, either in the area or list frame, the proration factor is one divided by the total number of partners, including those on the list and those not on the list. This is necessary because of the way the nonoverlap fraction was computed in the area frame. The nonoverlap fraction plus the proration fractions must equal one. The combination of the two frames will then properly represent the partnership.

Maximum Overlap Procedure (Alternative II)

This procedure differs from the partial only in the procedures for partnership operations after the data are collected. The decision diagrams are depicted in Figure II.

Alternative II is exactly the same as the partial procedure if a partnership or corporation has a name. If an area frame tract is a partnership operation, the partnership has a name and the name is on the list frame, the tract is overlap. The procedures differ when a corporation or partnership does not have a name. If any partner's name is on the list, the tract is overlap under Alternative II. Only if none of the names of partnership members are on the list does the tract become nonoverlap.

The procedure used to prorate partnership data reported by an individual whose name was selected from the list frame must be compatible with the procedure used to determine if the partnership is overlap when the individual is found in an area frame segment. Under Alternative II the proration factor applied to list data is one divided by the number of partners who are actually on the list frame, rather than the number of all partners in the partnership. If the partnership operation does not have a name, but an individual selected from the list reports for a partnership or corporation, the procedure is to prorate the data based on the number of times that operation could have been selected from the list. This requires survey statisticians to search their list for the name of

every partner reported on the list questionnaire.

Both procedures described so far are theoretically correct. They differ somewhat in approach and in problems of implementation. Alternative II requires all names associated with a partnership reported by an individual selected from the list frame be checked against the list. This does involve extra work; however, it provides an additional check for duplication in the list. If there is a tendency for names of partners that are not valid partners to be reported on a list frame questionnaire, proration effects are minimized by counting only those on the list. The current procedure does not require all names of partners reported in the list sample to be checked against the list. From the standpoint of editing time required, the partial procedure is an improvement. However, it is risky because the assumption that all names are those of valid partners may be seriously violated.

Exact Match Procedure (Alternative III)

This procedure is depicted in Figure III. It relies on the following basic assumptions.

a. An individual's name on the list represents a unique land operation only associated with that name. More specifically, the name Sam Jones can only represent land operated solely by Sam Jones. It cannot represent land operated jointly by Sam Jones and others.

b. If the individual does not have a unique land operation, all reported livestock data will be edited to zero.

The procedure is also theoretically correct. At the outset, it appears to be somewhat easier to understand and implement than the previous procedures. It will increase the size of the nonoverlap domain, however. The procedure also requires a much stiffer correspondence between the name of the operation and the name appearing on the list frame. For example, to identify the nonoverlap domain for partnership and corporation tracts, the procedure is as follows: If the partnership or corporation has a name, and the name is on the list frame, then the tract is overlap. If a partnership or corporation does not have a name, the tract is automatically nonoverlap.(Individual names listed together would constitute a partnership name.)

Under certain conditions the editing procedure for list frame sample units using Alternative III will differ from the other alternatives. If a partnership or corporation name is selected from the list, only reported data associated with that partnership are to be summarized. This does not differ from the previous two procedures. In fact the nonoverlap and the list frame procedures for partnerships are the same as for individuals for all three alternative methods if the partnership or corporation has a name. However, under Alternative III if a list frame sample unit is an individual name and some or all of the reported data are associated with a partnership or corporation, all partnership data would be edited to zero. There is no proration of partnership data under Alternative III. This means that all partnership or corporation data must be represented entirely by a single list frame sampling unit (Name) or the partnership will be entirely nonoverlap.

Comparison of Survey Results from Alternative Procedures

Although each nonoverlap procedure being discussed is theoretically correct, survey estimates and associated sampling errors can vary among procedures for any given survey. Table 1 shows the percent deviation in the survey estimate of total cattle comparing the alternative procedures to the partial procedure for each of six states analyzed. Shown in Table 2 are the relative sampling errors for each procedure.

Survey estimates of total cattle on a state-by-state basis did not vary substantially among the different procedures. Comparisons of six state totals for cattle show the three alternative procedures provide very similar results. This suggests that not only in theory but also in practice these different procedures can be expected to yield very similar results on a regional basis and more certainly on a national basis. Relative sampling errors associated with Alternative III were consistently larger than those associated with the other two procedures.

Table 1: Relative Size of Cattle Estimates

	States						
Procedure	1	2	3	4	5	6	Total
	%	%	%	%	%	%	%
Max. Overlap	+.7	- .7	+.5	+.8	-1.9	+ .3	-.3
Partial Overlap	-	-	-	-	-	-	-
Exact Match	-.7	+1.3	+.7	-.7	+ .7	+1.2	+.6

Table 2: Relative Sampling Errors in Cattle Estimates

	States						
Procedure	1	2	3	4	5	6	Total
Max. Overlap	3.5	3.3	3.6	3.4	3.6	6.7	1.6
Partial Overlap	3.6	3.3	3.6	3.5	3.8	6.7	1.7
Exact Match	3.9	3.7	3.8	4.0	4.4	6.9	1.8

Nonsampling Errors

The nonsampling errors associated with any procedure for determining overlap between the list and area frames are problems connected with associating a name with a unit of land. As the survey questionnaires for these six states were being re-edited, nine sources of nonsampling errors were observed in domain determination under the partial nonoverlap procedure. They are:

1. Incorrect proration of data among partners,
2. Partnership sampled but data are still prorated,
3. Partnership sampled but more partners are involved,
4. Livestock on individual and partnership land edited incorrectly,
5. Interpretation of arrangements sometimes differ between frames,
6. Not enough information available,
7. Reporting unit different from sampling unit,
8. Inconsistency in respondent reporting,
9. Failure to link names.

Potential errors 1-4 are specifically the result of the prorating activity done under the partial overlap procedure. These problems are less likely to occur under Alternative II and almost certainly will not occur under Alternative III. Errors 5-9 could be worse under Alternative III since all the data could be included or excluded incorrectly. However, with increased emphasis on the sample name and type of operation involved, additional information should be obtained before accepting or rejecting all the data. The relative size of the nonsampling errors discovered in the partial procedure are presented in Table 3.

Table 3: Nonsampling Error in Cattle Estimates

	States						
Domain	1	2	3	4	5	6	Total
	%	%	%	%	%	%	%
List	+.6	+ .2	+2.5	+ 1.9	+.2	+ .8	+ .9
Non-overlap	+.2	+11.1	-3.0	-16.3	+.3	-10.2	-1.6
Total	+.6	+ 1.4	+1.5	- 1.0	+.2	- 2.2	+ .5

A consistent upward bias was observed for list frame data stemming from a reluctance to edit out reported data even under strict decision rules. Approximately 15 percent of the estimate is derived from the nonoverlap domain and substantial differences sometimes occur here even when only a few reports were incorrectly edited. The nonsampling error over all six states was not serious.

Summary

An analysis of survey data from six states was used to compare three alternative procedures for determining overlap between the area and list frames.

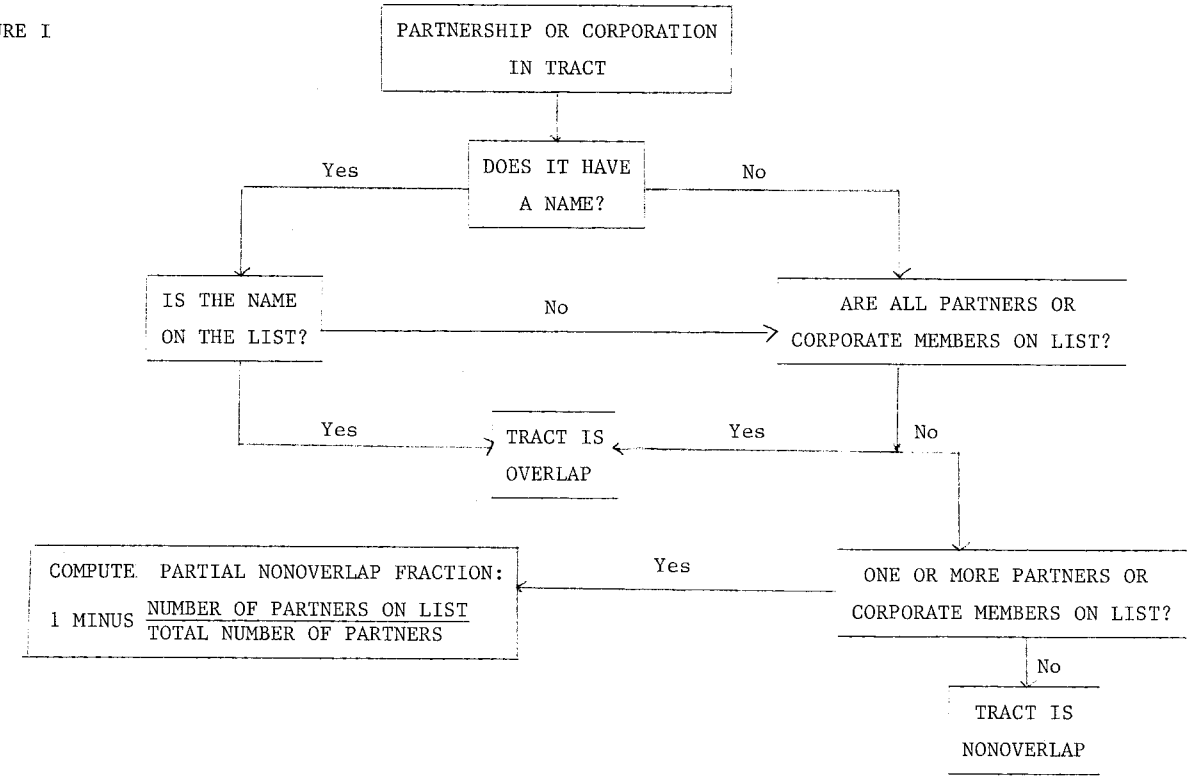
Comparisons of cattle estimates on a state by state basis are not inconsistent with the theoretical result. Therefore, the choice among the alternative porcedures should be based on ease of data collection and degree of nonsampling error. Experience gained through this research suggests Alternative III is easiest to apply and may be less susceptible to nonsampling error; however, the data suggest its use may result in slightly larger sampling error.

Footnote

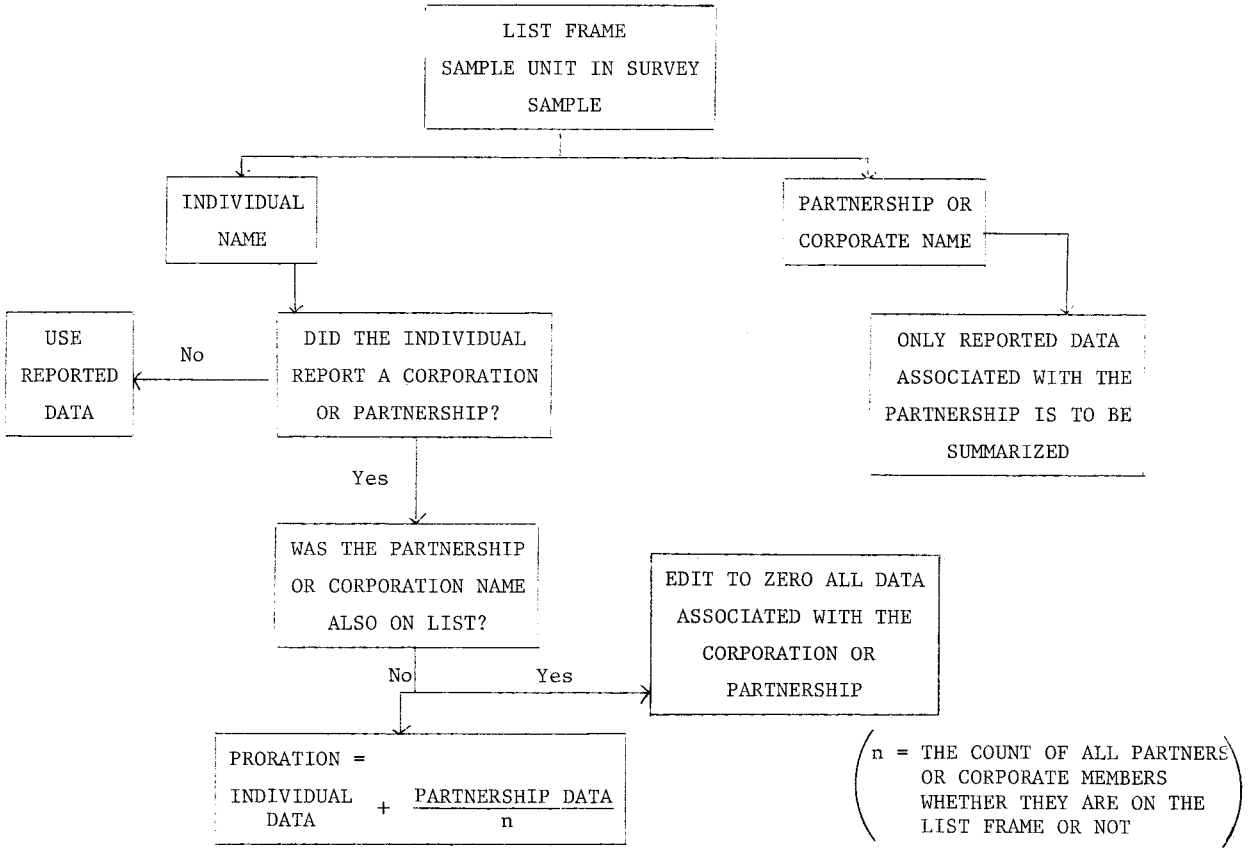
*Extracted from "Multiple Frame Livestock Surveys-An Evaluation of Alternative Methods of Overlap Determination" by Frederic A. Vogel, Raymond R. Bosecker, and Dwight A. Rockwell, Statistical Research Division, ESCS, USDA, June 1976.

A. Current decision diagram for partial nonoverlap concept applied to area frame.

FIGURE I

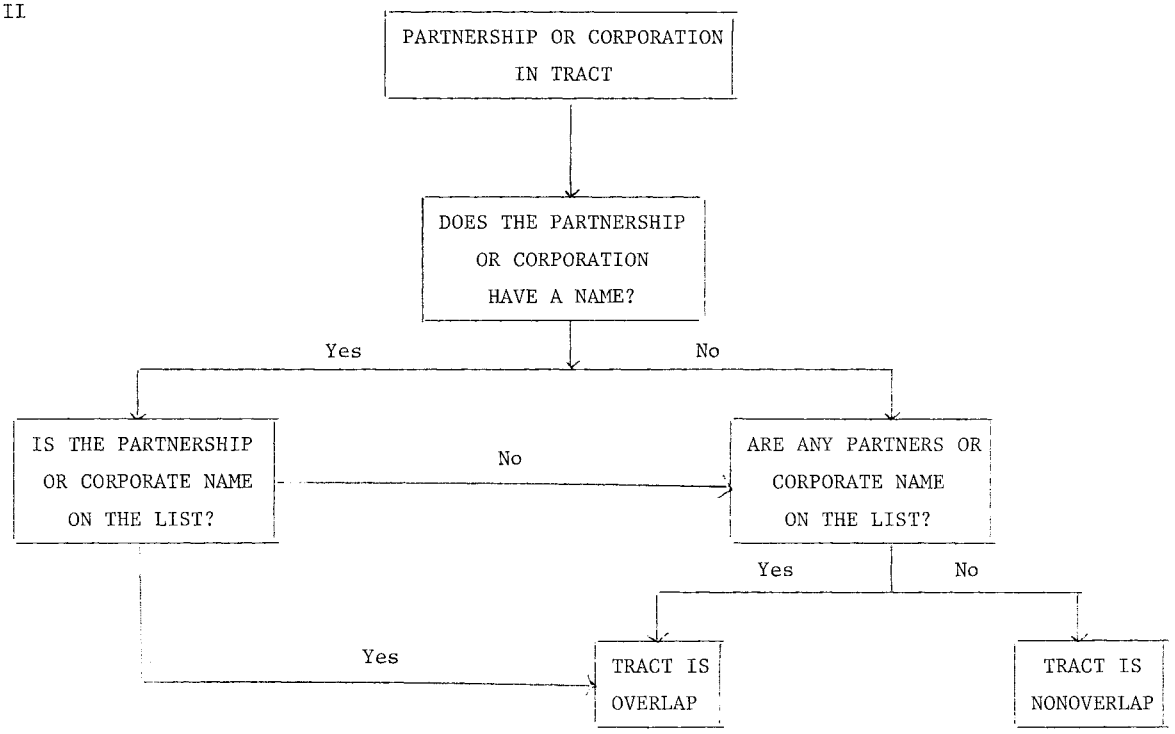


B. Current decision diagram for partial nonoverlap concept applied to list frame.

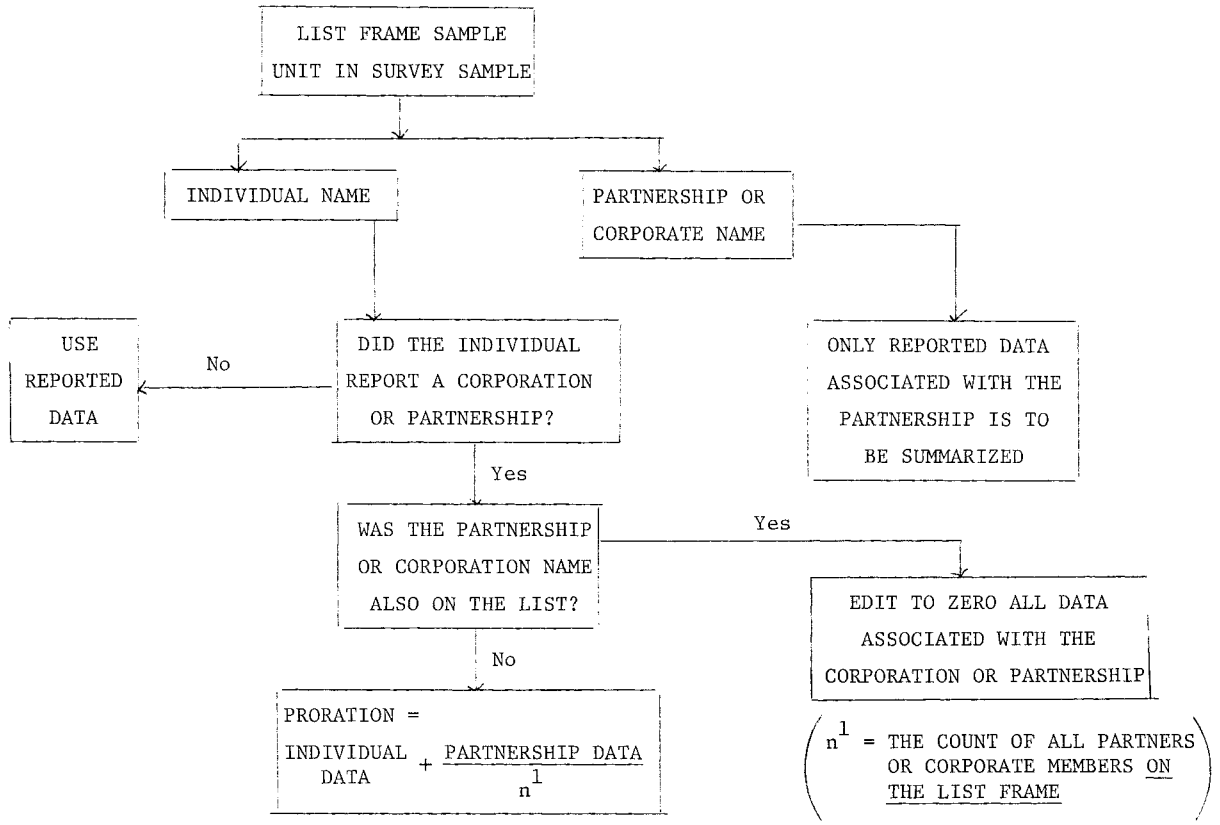


A. Decision diagram for maximum overlap concept applied to area frame.

FIGURE II

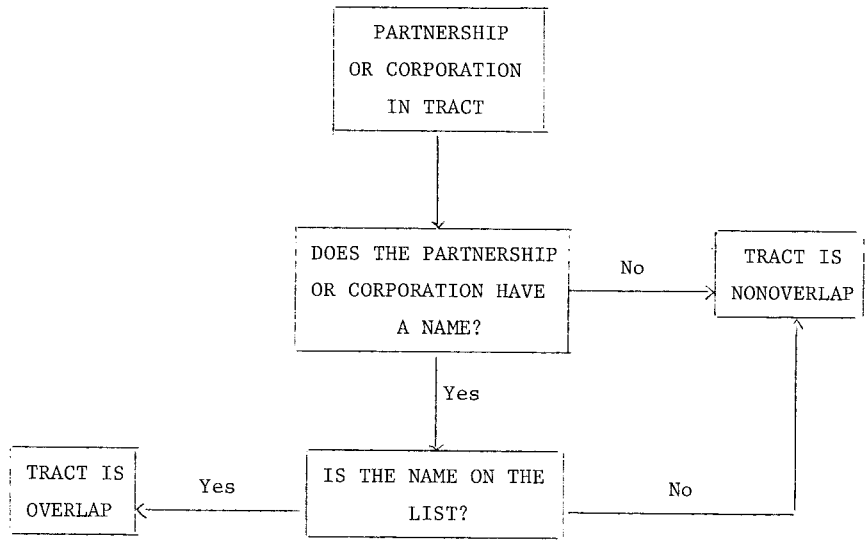


B. Decision diagram for maximum overlap concept applied to list frame.



A. Decision diagram for exact match procedure applied to area frame.

FIGURE III



B. Decision diagram for exact match procedure applied to the list frame.