METHODOLOGICAL DESIGN TO

ESTIMATE THE TARGET POPULATION

FOR THE MIGRANT HEALTH PROGRAM

1983

FINAL REPORT

SUBMITTED TO

THE OFFICE OF MIGRANT HEALTH
HEALTH RESOURCES ADMINISTRATION
ROCKVILLE, MARYLAND
CONTRACT NUMBER HRSA 83-24(P)

SEPTEMBER, 1983

SUBMITTED BY

ALICE C. LARSON, Ph.D.
7045 BEACH DR., S.W.
SEATTLE, WASHINGTON 98136

TABLE OF CONTENTS

Wallawla
Central Wash HSA (Popenich)

Alice

I.   CONTENT AND STRUCTURE OF THE FINAL REPORT

This Final Report describes the process, progress and conclusions of the development of a "Methodological Design to Estimate the Target Population for the Migrant Health Program, 1983." Work on this methodology and the conclusions reported here were undertaken by Alice C. Larson, Ph.D., who has been providing expert consultation to the Office of Migrant Health (OMH) in an advisory capacity on this subject since September, 1982. The methodology described continues a process undertaken by the OMH to define its target population, as reported in the 1973 and 1978 Migrant Health Program Target Population Estimates studies.

Original plans for this Final Report were to adhere to the four-phased methodological outline proposed by the consultant. However, the meeting of experts held at the Parklawn Building in July, 1983 has suggested a somewhat altered methodology which may not correspond to the established phases. To avoid confusion, this Final Report will continue to address each of the four original steps, discussing plans for the altered methodology under Phase III, as noted below.

* Phase I: Identify counties which employ seasonal farm labor -- discussion
               of the report prepared by Whittaker Medicus, Inc., under the
               guidance of the consultant, which presents data for 12 sample
               states.

* Phase II: Divide counties in which seasonal agricultural workers are employed
               into high impact, low impact or unclassified areas -- discussion
               of the sources and system used to transform and compare existing
               information on migrant and seasonal farmworker (MSFW) peak presence
               in each county, presentation of the resulting data and classification
               of counties in 12 sample states.

* Phase III: Categorize each unclassified county as a high or low impact
               area -- summary of the July, 1983 meeting of data experts which
               suggested a different approach to county classification and
               description of how this new approach should be applied.

* Phase IV: Verify the classification of counties as high or low impact
               through review by local sources -- detailing of the steps
               necessary to enact a local review process.

* Phase V: Other considerations in design of this.methodology -- discussion
of future work needed to adequately describe the Migrant
Health Program target population.

II. PHASE I

IDENTIFY COUNTIES WHICH EMPLOY SEASONAL FARM LABOR

The identification of counties which use seasonal agricultural workers was accomplished through a display of data from three existing sources: Employment and Training Administration, Form 223 (ETA 223), Census of Agriculture (COA) and the Migrant Education Program, Migrant Student Record Transfer System (MSRTS) enrollment figures. (These sources will be described further in the next section of this Final Report.) Information from each source was gathered for every county in the 12 sample states. The data were obtained and arrayed by Whittaker Medicus, Inc., a technical assistance contractor of the OMH. These numbers are presented in Medicus' February, 1983 report, "An Enumeration of Migrant and Seasonal Farmworkers in Twelve Selected States." The consultant guided the gathering of this information.

Contact names and information request procedures were provided to the contractor for obtaining data sources. Some of this information was more difficult to secure than originally thought. COA information was readily available in a Bureau of the Census publication, although data pertained to 1978. A more recent COA survey was completed in 1982, but this information has not yet been published and is not available from Census prior to publication.

MSRTS numbers were first sought from the Migrant Education Program in Washington, D.C., but the contractor was referred by this office to the central MSRTS data bank in Little Rock, Arkansas. Numbers were received for every state, but the county information was coded. The Little Rock office did not have code breakdowns to county names making it necessary to contact the Migrant Education Office in each of the 12 states. Securing MSRTS enrollment information was a time-consuming pro-cess, but a pattern for obtaining these numbers was set which can be followed if MSRTS numbers are needed for other states. It also seems possible to secure

enrollment figures directly from each state Migrant Education Office. In gathering

this information from any source, it is important to be sure the figures are

unduplicated enrollment counts not migrant full-time-equivalency numbers.

The most difficult information to secure was ETA 223. The Washington, D.C.

office of the U.S. Department of Labor (DOL) has summary information sent by state

departments of employment, but these numbers were found to be incomplete and only

available by Agriculture Reporting Area, a DOL designation which groups several

counties based on employment needs and agriculture activity. It was necessary

to contact each state department of employment to obtain the necessary information.

No contact names were available in the states, and at first, it was very difficult

to find the correct individual and have the party understand what information was

needed. Several guidelines developed which seemed to prove successful:

1. Call the department of employment main office in the state capital and start with either a "research and statistics" office or a division concerned with "agricultural labor."

2. Insist on talking to the person who prepares "ETA Form 223 -- In-Season Farm Labor Report." For most states, this will be an individual who coordinates ETA 223 reports from local employment offices. The best rule to follow is the "right person" will know the exact form in question. Often, researchers were referred to the "Monitor Advocate" who also works with MSFWs but who does not complete the ETA 223 form and is not the person needed. In addition, the information desired is not how many MSFWs "registered for work" at each local employment office.

3. When the correct person is found, ask him/her for information on the number of hired seasonal (sometimes called "local") and migrant (usually divided into "interstate" and "intrastate" and/or called "nonlocal") workers reported for every county in the state for each of the last 12 months or for the months of heaviest agricultural activity. Some states may publish this information, but most will have to be asked to assemble the numbers.

4. It is important to stress that information is needed on both migrant workers and seasonal workers. Data should be by county or by local reporting office (a small group of counties), but it needs to be broken down smaller than an ARA. It also seems important to stress that the researcher realizes the MSFW numbers supplied are only an estimate, that they may or may not reflect an accurate picture and that real numbers on MSFWs are very hard to obtain.

5. Document the general process used to collect the number; e.g., local office

personnel make estimates they report to the state, one dominant agricultural industry is asked to report monthly figures, a formal survey of growers is taken during certain months, etc. It is not necessary to have exact details, but it is helpful in understanding the strengths and weaknesses of the collected information to know the general system used to gather numbers.

In summary, ETA 223 data take persistence, persuasion and time to collect. The major problem in producing the Phase I report was waiting for states to send ETA 223 information and calling state personnel if the wrong information was received. Now that this pattern to secure ETA 223 data has been established, it should be somewhat easier to gather these numbers, but it will still take time as it involves contacting individual state agents.

Collected information from ETA 223, COA and MSRTS were displayed in a column arrangement by state and county in the Medicus report. This seems to be a satisfactory data array as long as unusual features of any of the pieces are clearly identified through footnotes. As indicated above, weaknesses, strengths and methods used by local sources to collect this information (particularly ETA 223) should be documented.

III. PHASE II

DIVIDE COUNTIES INTO HIGH IMPACT, LOW IMPACT AND UNCLASSIFIED AREAS

DATA RESULTS FOR TWELVE SAMPLE STATES

A. Introduction

This section presents information on "Phase II" of the original four-phased methodology to estimate the target population for the Migrant Health Program, 1983. The purpose of this Phase was to divide counties in which seasonal agricultural workers are employed into one of three categories: high impact, low impact or unclassified. For purposes of Phase II, the definition of each (category) was as follows: high impact -- 5,000 or more MSFWs, including workers and dependents, at peak presence; low impact -- under 3,000 MSFWs present at peak; and unclassified -- between 3,000 and 4,999 MSFWs at peak.

The method used to perform the designated task was to compare numbers of MSFWs from existing data sources to place counties in their appropriate category. Guidelines were also established to cover situations where the sources disagreed on the number of MSFWs. Information was gathered on ten sample states: Colorado, Connecticut, Florida, Georgia, Michigan, New Mexico, North Carolina, Ohio, Texas and Washington. Information was readily available for two additional states, Idaho and Oregon, resulting in their inclusion as well.

The narrative is divided into three parts: (1) Discussion of the Data and Directions for Use, (2) Data Summary, and (3) State Tables: "Discussion of the Data" includes: "Sources for Numbers," "Calculation of Numbers," and "Instructions for Use." "The Data Summary" presents a listing by state of every high impact or unclassified county, and maps for the states and existing Migrant Health Clinic service areas. The State Tables display the actual calculated numbers derived from each data source and the appropriate category for each number.

B. Discussion of the Data and Directions for Use

1. Sources for Numbers

a. Primary Data Sources

Two sources of numbers were used for all twelve states. These are information from the Employment and Training Administration Form 223 (ETA 223), "In-Season Farm-Labor-Report," as supplied by each state department of employment in accordance with requirements of the U.S. Department of Labor; and the Census of Agriculture (COA), conducted by the Bureau of the Census, U.S. Department of Commerce.

ETA-223 data are accumulated yearly for each month of concentrated agricultural activity in which seasonal farm laborers are employed. This period in each state may differ but usually covers from April through September. Many states report information for all 12 months. Original numbers used in this Report are for peak employed MSFWs for 1980-1981. These were obtained from each state employment division staffer in charge of preparing ETA 223 forms.

COA data are taken from a mail survey of local growers conducted every five years by the Bureau of the Census. This report uses 1978 figures, the latest available, for every state and county.

These two data sources do not necessarily provide the best information useful to the needs of the OMH. However, these are the only sources which make estimates of the MSFW population by county, on a systematic basis, available for every state in the country. It is important, therefore, to be clear on the weaknesses of these sources, the methods used to gather information and the relationship between MSFWs reported and MSFWs included as the target population for the Migrant Health Program.

Both ETA 223 and COA data relate to employed seasonal laborers but cover different periods of time and are collected in different manners. ETA 223 information

covers migrant and seasonal workers employed in farm labor on the specific day

of the monthly count. These figures cannot be added to form a yearly employed MSFW

figure because of duplication of individual workers in monthly counts. COA data are

reports of agricultural workers hired at any time of the year by individual employers

(growers). These are yearly employment figures, not peak totals, and they contain

duplicate worker counts, as MSFWs may work for more than one local grower.

The DOL formal definitions for the data collected on the ETA 223 form are as

follows:

Seasonal Workers are those hired or assigned to work on any farm or establishment
for less than a continuous 150-day period in the course of one year. They may
be local or migratory workers.

Local Workers are those who regularly reside within normal daily commuting
distance of their employment.

Interstate Migratory orkers are farm workers whose place of normal residence
is outside of the state in which work is located.

Intrastate Migratory Workers are farm workers whose normal living quarters
are elsewhere in the state but who reside temporarily within the locality of
employment for purposes of engaging in seasonal farm work.

A more informal definition seems to be used in actual information gathering.

DOL allows each state to gather numbers of MSFWs using the best available method.

The agency does not specify the process. This means that each state uses a different

and often very unscientific methodology. California may have one of the most

sophisticated data collection systems, using a computer to estimate employed seasonal

labor. Many states rely on the services of one expert in the local Job Services

office who one day a month drives within his designated area and counts workers in

fields, talks to growers, visits labor camps and generally just looks around. From

this information, he/she makes an exact estimate of employed seasonal workers.

Some personnel indicate they determine which seasonal workers are local and which

are migratory by "counting the out-of-state cars." Connecticut uses a different

system -- asking the tobacco industry to supply the monthly figures depending on the number of seasonal workers employed in tobacco crops.

Additional problems with ETA 223 information include missing many MSFWs who are not employed on the date of the count, e.g., the harvest may have ended the day-before; not counting MSFWs employed in related agricultural production activities; and including college and high school students, housewives and others who are not considered MSFWs under the Migrant Health definition. (InterAmerica Research Associates, An Evaluation of Access To and Availability of Human Resources Programs for Migrant and Seasonal Farmworkers in Region X, performed under contract to the Office of Planning and Evaluation, HEW, Region X, Seattle, Washington, 1978, pp 223-224.)

COA survey forms ask growers to report two categories of workers: those employed under 150 days per year and those employed 150 days or longer. No further definition is required for this data source. Workers identified as employed under 150 days can be said to be seasonal laborers, although there is no breakdown to "local" or migratory workers, as with ETA 223 information. Problems with COA data include: count duplication from grower to grower; counting anyone hired by the grower (e.g., family members, as well as others whose primary occupation is not farm work); and including cattle, poultry and dairy workers in the counts (these are excluded in the Migrant Health definition). Additionally, the information is gathered through a self-reporting survey, which may present other problems; e.g., non-reporting, skewed sample, and misreporting.

b. Secondary Data Sources

Information from data sources in addition to ETA 223 and COA is presented for Michigan, Ohio and Texas. This information was gathered by a state government agency. Additionally, numbers from the Migrant Education Program, MSRTS are

documented for Texas. The MSRTS keeps information on the number of migrant students enrolled in local Migrant Education Programs. This can be broken into county data. The MSRTS was considered for use in every state, but it was found to vary too widely in accuracy, as the systems local numbers depend on the ability of the local Program recruiter to enroll students. Also, some individual school districts in which migrants are present may choose not to have a Migrant Education Program and, therefore, report no numbers. MSRTS information is given for Texas as it was found that ETA 223 and COA information reports only working MSFWs, excluding many migrants who do not perform seasonal farm work where they live but migrate throughout the United States.

The Michigan source is an estimate of migrants living in housing. This information was submitted by the State of Michigan, Department of Labor, Michigan Employment Security Commission and reflects the 1982 agricultural season.

The Ohio Source is a 1981 census of migrants conducted by the Ohio Bureau of Employment Services, combined with in-the-file information on seasonal workers. Numbers were obtained from local employment office records, welfare and Food Stamp Offices, migrant programs -- including migrant health clinics, and education programs. The definition of migrant used in the study was:

A seasonal farmworker who had to travel to do farmwork so that he/she was unable to return to his/her permanent residence within the same day, working at least on aggregate of 25 or more days or partial days performing farmwork, earning at least half of his/her monies from farmwork and was not employed in farmwork year-around by the same employer.

The report on the Ohio study indicated that 1981 figures showed a decrease of 19% from 1980 figures due to "use of mechanical harvesters, decreased acreage and a season of inclement weather conditions." It also indicated that the tomato industry expects to add acreage in tomatoes which should increase the need for migrants in the future.

The second additional Texas source was also a state-wide survey conducted by the Governor's Office of Migrant Affairs (GOMA). This study, conducted in 1976, used a specially designed methodology which combined interviews of MSFWs with an estimation of exact numbers based on the sample surveyed. The process used was to locate MSFW clients of the Texas Employment Commission and apply a snowball technique having them identify friends and acquaintances who "are involved in similar occupations" who were then interviewed. The definitions used in this study were:

   Seasonal farmworkers: those involved in farm work for five months or less yearly and who did not leave their residence to obtain agricultural type work.

   Migrant farmworkers: same as seasonal farmworker but did leave their residence to obtain agricultural type work.

MSRTS data used as the second additional source for Texas were obtained from the central office in Little Rock, Arkansas. This information was compiled from the computerized records of the MSRTS and included an unduplicated count of students enrolled in Texas Migrant Education Programs for the school year 1981-82.

2. Calculation of Numbers

   To compare ETA 223 and COA figures they each had to be adjusted to reflect MSFWs (including dependents) at peak presence. Neither source, in its original form, provided this information as they report only employed workers. In addition, COA figures gave yearly totals not peak presence. It was necessary to derive factors to account for dependents per worker, to adjust COA numbers to peak presence, and to apply an available error rate to ETA 223 information. Every derived number was rounded up, reflecting the reality that these figures relate to people.

   a. ETA 223

Dependent factors used with ETA 223 data were available from the 1973 Migrant

Health Program Target Population Estimates report which used ETA figures as its
primary source. Factors gave an average number of persons per worker for migrant
and for seasonal households in each state. These were based on state or regional
studies available at the time of the Report. The 1978 Migrant Health Program
Target Population Estimates report used these same factors. Although dependent
estimates are over ten years old, no better estimates are available for all the states.
Therefore, use of these dated factors denotes uniformity with past practice, but may
reflect inaccuracies. The dependent factors often varied throughout the 12 states
in this report.

Both the 1973 and 1978 Target Population reports estimated a data error rate
to account for MSFWs missed in the counts, data collection errors and other factors.
The error rate reported in the 1973 report was 25%, and in the 1978 report it was
33%. For this document, an error rate equal to the average of these earlier
reported rates, 29%, was used for ETA 223 information from every state. It is
unknown whether this rate accurately reflects the error in the original information,
although, due to cutbacks in state funds available at the local level to gather
such statistics, there may be reason to believe that accuracy of estimation has
declined and 29% may underestimate the true error rate.

Calculation of final figures for ETA 223 data, presented in the state tables,
used the formula:

1. Migrant number X migrant dependent factor + seasonal number X seasonal
   dependent factor = MSFWs (including dependents).

2. MSFWs X 29% error rate + MSFWs = MSFWs adjusted for data errors.

Original ETA 223 information in some counties did not provide separate migrant
and seasonal worker numbers but only a combined MSFW employed laborer total. For
these areas, an average of the migrant and seasonal dependent factors was applied
and the result increased by the 29% error rate. In several states no separate

county figures were available but only a total for two or more counties which reported data to one local office or for a group of counties identified as an ARA. For these states, it was assumed that the percent distribution of workers in the sum of the counties on ETA 223 was the same as the percent distribution of combined COA data which were available for each county separately.

Colorado, Connecticut, Florida and Texas had ETA 223 information grouped in from 2 to 14 counties, although there were several single county reports in these states. Georgia and Michigan had ETA data grouped by ARA ranging from 8 to 25 counties. Idaho, New Mexico, North Carolina, Ohio, Oregon and Washington had separate county data in their ETA reports.

It was not necessary to adjust ETA 223 information to reflect peak numbers as this source reports monthly totals. It was assumed that these reports reflected maximum employment for the month (which may or may not be accurate) and the largest monthly figure was, therefore, taken to represent peak employment.

b. Census of Agriculture

COA information is presented in two categories: workers hired less than 150 days and workers employed 150 days or more, with the former category equating roughly to seasonal agricultural workers. Definitional problems with COA information and possible count duplication were discussed in the "Sources for Numbers" part of this section. It is not possible, however, to determine how much these problems affect the data and no other Office of Migrant Health report has used this source. For these reasons, no error rate can be determined and, consequently, none is applied to COA figures. It was possible to adjust these figures to include dependents using the same factors applied to ETA 223.

The major problem in adjusting COA data was converting year-around employment figures to peak presence. Dick Storm from the Bureau of the Census' branch in

charge of the COA, Robert Coltrane from the Rural Labor Market Section of the

U.S. Department of Agriculture (which collects and tabulates MSFW estimates) and

agricultural extension agents attached to universities in Oregon and Washington

could not suggest an appropriate means to adjust year-around figures. The

relationships among the number and type of crops grown in counties, the length of

the agricultural season and the COA numbers were examined but no clear pattern

emerged. Lacking any other logical adjuster, it was decided that the ratio of

workers employed under 150 days to those employed 150 days or more related to

peak presence for seasonal laborers (those working under 150 days). Because

this relationship was unproven, two separate estimates were made based on different

adjustments. The result supplied two COA estimates of MSFWs per county which

could be compared to ETA 223 county calculated numbers.

COA published data presents information in four columns: (1) the number of

workers reported by farmers indicating they hired only those working less than

150 days; (2) the number of workers reported by farmers who hired only workers em-

ployed 150 days or more; (3) for farmers hiring both types of workers, the number

they reported employing less than 150 days; and (4) for farmers hiring

both types, the number they indicated worked 150 days or more. COA peak presence

adjustment Method #1 used all four columns of numbers to derive three ratios which

were then averaged. The steps in this method are noted on the following page in

reference to the data example.

	Column 1 Report Only Workers Less Than 150 Days	Column 2 Report Only Workers 150 Days Or More	Report Both Types of Workers	
			Column 3 Work Less Than 150 Days	Column 4 Work 150 Days Or More
[Example]	2,743	149	827	414	]

Step 1: Column 1 + Column 3 ÷ Column 1 + Column 2 + Column 3 + Column 4
[(2,743 + 827)] ÷ (2,743 + 149 + 827 + 414) = .864]

Step 2: Column 1 ÷ Column 1 + Column 2
[2,743 ÷ (2,743 + 149) = .949]

Step 3: Column 3 ÷ Column 3 + Column 4
[827 ÷ (827 + 414) = .666]

Step 4: Mean of Steps "a", "b", and "c"
[(.864 + .949 + .666) ÷ 3 = .826]

COA Peak presence adjustment Method #2 provided a ratio which considers the need for only seasonal laborers, which reflects the demand for hiring workers in peak periods. It develops a ratio based on the number of workers reported by growers hiring only seasonal laborers to growers hiring both types of laborers. Using the column headings and example for Method #1, this method is calculated as follows:

Column 1 ÷ Column 1 + Column 3
[2,743 ÷ (2,743 + 827) = .768]

Factors based on the two calculation methods were developed for every ARA in each state. Factors were applied to the counties which made up the ARA. For each of the 12 states, Method #1 resulted in a larger factor than Method #2 with the exception of four separate counties in New Mexico and two ARAs in Texas. The difference in the factors calculated by the two methods varied from .001 to .677 for all the ARAs.

- 16 -

c. Other Data Sources

The additional data source for Michigan provided an estimate of migrant workers as supplied by the Michigan Department of Labor. These figures were multiplied by 2.2, the factor from the 1973 PHS Report for dependents. The resulting figures were taken to represent peak numbers for migrants only.

The Ohio source which combined Ohio Bureau of Employment Services information with a migrant survey presents data on MSFWs. Migrant numbers were given for "workers [age] 14 and up" and "non-workers." The results roughly equal migrants and dependents. Data for seasonals were only provided for workers; therefore a factor of 2.78, as taken from the 1973 PHS Report, was used to convert these numbers to include dependents. The migrant figures only account for interstate migrants, and both migrant and seasonal numbers are assumed to represent peak presence. It is also interesting to note that the migrant family figures presented in this survey indicate the average number of people in the household per worker as 1.56 (1.58 excluding single person households -- single migrants without dependents made up only 3% of the work force). These figures indicate a shift from those in the 1973 PHS Report which showed 2.2 persons in the household per migrant worker. The 1981 dependent figure was not used in calculation of ETA 223 and COA as no update for the 1973 Report figures was available for the other 11 states.

Numbers form the Texas GOMA study covered both migrants and seasonals including their dependents. Figures were collected between September and April, and the report notes "Those potential respondents not at their home base during this period were excluded." This study provided information by residence and does not indicate that MSFWs actually worked in any county. Collection of this residence data at the time of year most MSFWs would not be employed in farm work equates to the peak presence data from the ETA 223 and COA sources. GOMA figures, in fact, compliment numbers calculated from the other two sources as ETA 223 and COA data

refer to employed workers while GOMA reports the residence of such workers. As discussed in the "Sources for Numbers" section, this is important as many MSFWs in Texas seem to work in a different county or state than the one in which they reside.

MSRTS figures, used for Texas, had to be adjusted in several ways. The original numbers roughly cover only migrant children ages 5 to 17 enrolled in a local Migrant Education Program. The definition of "migrant" includes children whose families are currently migratory as well as those whose families have not migrated for up to five years. Technically, settled-out migrant children whose families are not mobile do not need to have family members who are seasonal farmworkers. In reality, however, most children in the Program are mobile migrants. MRSTS data must be expanded by factors which include an error rate for number of migrant children ages 5 to 17 not enrolled on the MSRTS, the number of children ages birth to five years and the number of adults. All figures must be based on available MSRTS data and must be calculated per family, requiring a figure for the number of children per family in the Program. These calculations automatically exclude single persons or childless families and those with only children younger than five years. In addition, MSRTS figures are reported for year-around not peak presence, however, if the logic of the GOMA report is used, it can be assumed that because Texas is a primary migrant home base, year-around figures approximate peak presence.

Through the assistance of personnel in the Migrant Education Program, contact was made with Mike Hoffman of the Indiana Department of Public Instruction, Division of Migrant and Bilingual/Bicultural Education, who under contract to the Migrant Education Program conducted an intensive study of migrant children not enrolled in the MSRTS. Mr. Hoffman collected information from a sample of migrant households in Indiana which indicated 30% of children ages 5 to 17 were not enrolled in the MSRTS, and the average number of migrant children in the system per household

was 3.2. Mr. Hoffman also found an average of two adults per migrant household.
He had no information on the number of children under age five per household.

A survey of patient records in a migrant health clinic in Washington indicated
an average of 1.143 children per migrant household under age 5, although other
household data did not concur with the Indiana survey; e.g., Washington patient
information showed 1.714 children ages 5 to 17 per household and 1.777 ages 18
or older per household. (Survey was of 1,227 persons identified as migrants and
their household members where age information was available -- 59% of all migrant
patients -- from the Walla Walla Community Health Clinic patient ledgers, Walla Walla,
Washington, April, 1983). Enrollment information obtained from migrant child care
programs in Idaho indicated 1.58 children under age 5 per household. This information
is admittedly more heavily weighted toward families with this age child, which
corresponds to the age for which child care is provided. The Idaho information
also differed from the Indiana study by indicating 1.57 children ages 5 to 17
per household. No information was available on the number of adults per household.
(Client information is from the Idaho Migrant Council Head Start Program enrollment
for 428 persons ages birth to 17, 136 households, at the Burley, Twin Falls and
Weiser, Idaho child care centers, 1982.)

It was felt that the Indiana information on the number of children per
household ages 5 to 17 might relate better than the Washington and Idaho family
information to calculate multipliers for the Texas MSRTS data. Thus, it was
assumed that there are 2 adults and 3.2 children ages 5 to 17 per migrant household.
The 1.1 and 1.6 estimates of children under age 5 obtained from the clinic and
child care samples are probably weighted toward an overcount, as children in this
age group are a major proportion of patients seen at these two programs. Therefore,
it was felt reasonable to downgrade this estimate and assume 1.0 children per
household under age 5. When all of these factors are added, a total of 6.2 persons
per household (for those with children) is derived. This is below the 1973

PHS report estimate of 7.5 persons per household for Texas migrants and slightly above the 1976 GOMA study estimate of 5.84 persons per MSFW household.

Each Texas county's MSRTS figure was adjusted, using the factors mentioned above, as follows:

* Increased 30% to account for missed children ages 5 to 17

* The increased number divided by 3.2 to derive number of families.

* Number of families multiplied by 2 to determine adults.

* Number of families multiplied by 1 to determine children ages birth to 5 years.

* Number of children ages 5 to 17 (increased by 30%) added to number of adults and number of children under age 5 to obtain total number of migrants including dependents.

The resulting figures do not include seasonals or persons in migrant households without children. As indicated, MSRTS figures are decidedly weak because they exclude segments of the population and must be adjusted using several unproven factors. Therefore, MSRTS information was not used for every state. They are presented for Texas, however, as they provide an additional residence-based source to compliment ETA 223 and COA data.

d. Calculation of Overall Category

For all 12 states, the catagories (1 = high impact, 5,000 or more MSFWs; 2 = low impact, less than 3,000 MSFWs; 3 = unclassified, between 3,000 and 4,999 MSFWs) in which the calculated ETA 223 estimate and two COA estimates fell were summarized to derive an "Overall Category" to determine if the county was a high or low impact area or was unclassified. Rules were developed to guide this summary as many counties had data estimates which disagreed. In general, where there was any doubt as to high or low impact, the county was declared unclassified;

e.g., if one source classified as a high impact and one as a low impact, the county was said to be unclassified. Classification by two sources in the same category with the third source in disagreement derived an overall category in line with the former sources. The one exception was when the disagreement was between high and low impact, in which case the county was said to be unclassified.

Specifically, the rules governing county overall category classification were as follows:

1 = high impact

2 = low impact

3 = unclassified

$$1 + 1 + 1 = 1$$
$$1 + 1 + 3 = 1$$
$$1 + 1 + 2 = 3$$
$$1 + 2 + 2 = 3$$
$$1 + 2 + 3 = 3$$
$$2 + 2 + 2 = 2$$
$$2 + 3 + 3 = 3$$
$$2 + 2 + 3 = 2$$
$$3 + 3 + 3 = 3$$

The two additional data sources used in Michigan and Ohio did not result in a re-examination of each county's Overall Classification as it was felt that ETA 223 and COA sources in these two states were as reliable as in other states. However, in Texas, ETA 223 and COA represented employment-based information while GOMA and Migrant Education, reflected residence-based data. In this state an "Adjusted Overall Category" was necessary which considered all the data source estimates. As with earlier rules for Overall Category classification, if there was any disagreement between high and low impact the county was declared unclassified. In comparing figures calculated from the four sources (COA provided two estimates from one source), MSRTS numbers were probably an underestimate as they did not include all seasonal farmworkers. The two separate COA based estimates were considered as a unit in determining Adjusted Overall Category; i.e., if one COA

estimate disagreed with all the other sources, it would be disregarded.

The rules governing the adjusted Overall Category classification for each Texas county were as follows:

1 = high impact

2 = low impact

3 = unclassified

1 + 1 + 1 + 1 + 1 = 1
1 + 1 + 1 + 1 + 3 = 1
1 + 1 + 1 + 1 + 2 = 3
1 + (1 + 2 -- COA estimates) + 1 + 1 = 1
1 + 2 + 2 + 1 + 1 = 3
1 + 2 + 2 + 2 + 1 = 3
1 + 2 + 2 + 2 + 2 = 3
2 + 2 + 2 + 2 + 2 = 2
2 + 2 + 2 + 2 + 3 = 2
2 + (3 + 3 -- COA estimates) + 2 + 2 = 3
3 + 3 + 3 + 3 + 3 = 3

3.   Instructions for Use

     a. Regular State Tables

The following tables for each state present calculated figures for MSFWs at peak presence per county derived form ETA 223 and COA. Calculated MSFW figures are presented for additional sources, as described earlier, for Michigan, Ohio and Texas. Counties are divided into Agricultural Reporting Areas (ARA), a term used by the U.S. Department of Labor in ETA 223 data to denote "a geographic division within a state which is reasonably integrated in terms of farm labor market characteristics; has a supply of, or demand for, seasonal hired farmworkers; and where 50 or more seasonal hired farmworkers are employed at any time of the year." (Employment Security Manual, Part III, Section 4800.)

The following page shows an example of how each table's figures are to be read.

| County -- ARA | ETA 223 | | | | Census of Agriculture | | | | Over All Cat |
	Mig X 2.4	Seas X 2.94	MSFW X 29%	Cat	#1 X Varies X 2.67	#2 X Varies X 2.67	#1 Cat	#2 Cat	
Johnson	641	325	1,247	2	2,668	2,364	2	2	2
Washington	2,533	503	3,917	3	11,395	10,095	1	1	1
WESTERN	3,174	828	5,164	1	14,063 [X.464]	12,459 [X.411]	1	1	1

The first column contains the county name in lower case letters and the ARA in capital letters. The ARA is a summary of the county figures presented immediately above the ARA. The next four columns present ETA 223 data. The "Mig" column heading indicates the factor for dependents used to increase the original migrant worker ETA 223 number. For this example, the factor is "2.4." The "Seas" column heading, likewise, presents the dependent factor used on ETA 223 seasonal worker figures, "2.94" in the example. These household factors vary per state. The "MSFW" column heading indicates an error rate factor of 29% which was used to increase the calculated Mig and Seas figures. The "Cat" column under the ETA 223 heading indicates the category in which the calculated MSFW figure falls: Cat 1 = $5,000 \leq X$, Cat 2 = $X < 3,000$, Cat 3 = $3,000 \leq X < 5,000$.

COA data are presented in the next four columns. As noted in the "Calculation of Numbers" section, two separate MSFW estimates were derived from COA figures. These are identified in the table as "#1" and "#2." The "#1" column heading indicates the original source has been multiplied by a factor which "Varies." This factor has been calculated by ARA, as explained in the earlier section, and is listed in the table under the ARA summary figure for COA column #1. In the example, this factor is ".464" for the #1 method of calculating COA numbers. The #1 column heading also notes the original data were multiplied by another factor, "2.67" in the example, to include dependents, (this factor is an average of the "Mig" and "Seas" dependent factors used on ETA 223 figures.)

The next "#2" figures column presents calculated COA data using Method #2. Similar to the "#1" figures column, the heading indicates data were mulitplied by a factor which varies, and is presented below the ARA summary figure in column #2, and by a dependent factor. In the example, method #2 multiplied COA numbers for Johnson and Washington Counties by ".411" and by a dependent factor of "2.67". The next two columns indicate in which categories calculated COA figures fell; "#1 Cat" indicates the category for method #1, presented in the "#1" figures column, and "#2 Cat" shows the category for method #2.

The last column in the table, "Over All Category," shows the category in which each county and ARA were placed based on calculated ETA 223 and COA figures. The determination of this placement is governed by rules presented in the "Calculation of Numbers" section.

b. State Tables for Michigan, Ohio and Texas

Michigan, Ohio and Texas, tables present additional data sources as explained in the "Sources for Numbers" section. In Michigan and Ohio additional sources were not used to determine the "Over All Category" for the counties in the two states. "Adjusted Overall Category" is presented for Texas which incorporates numbers from ETA 223, COA and the two additional sources.

The Michigan table includes calculated figures from a state survey of migrant housing. The column heading on the Michigan table indicates this source under the title "Housing" and notes the original figures were multiplied by a factor of 2.2 to denote dependents per migrant worker. These calculated figures are for migrants only and do not compare to calculated MSFW numbers derived from the other sources.

The table for Ohio counties and ARAs include calculated figures from a survey of migrants and file data on seasonals, listed under the title "Survey." The original numbers for seasonal workers were multiplied by 2.78 to include dependents.

These calculation operations are indicated in the column heading. The final
numbers are for MSFWs and compare to MSFW figures derived from ETA 223 and COA
data.

Two additional sources are included in the table for counties and ARAs in
Texas. These are Migrant Education calculated figures for migrants ("Migrant
Ed" in the column heading), and figures from a study conducted by the Governor's
Office of Migrant Affairs ("GOMA" in the column heading). The section on "Cal-
culation of Numbers" indicates the method and rationale used to adjust Migrant Ed
original numbers to migrants at peak. The column heading indicates this operation:
the original number for migrant children enlarged by 30% to adjust for migrant
children missed by the program, and then increased for number of adults per family
("X" ± 3.2 children per family multiplied by 2 adults per family) and for number
of children ages birth to five per family ("X" ± 3.2 children per family multiplied
by 1 child under age five per family). The resulting calculated figure is only
for migrants, not MSFWs.

The "GOMA" column in the Texas table presents information taken directly
from the GOMA report. These figures have not been adjusted as they represent
MSFWs at peak and are comparable to ETA 223 and COA calculated figures presented
in the table. The final column, "Adjusted Over All Category," shows a recalculated
Over All Category considering data from the four sources: ETA 223, COA, Migrant
Ed and GOMA.

C. Data Summary

The following counties in each of the twelve states have been categorized as
either high impact or unclassified. The total for each state in these two categories
is also given. These state summaries are followed by state maps which indicate
high, low or unclassified counties and maps which categorize existing Migrant

Health Clinic service areas as high, low or unclassified impact. Counties which

make up service areas for clincs in the twelve states were taken from the Migrant

Health Directory, Public Health Service, Health Resources and Services Administration,

Bureau of Health Care Delivery and Assistance, Rockville, Maryland, January, 1983.

_1. Counties Categorized as High or Unclassified Impact

COLORADO

High Impact
    Weld
    Total Counties: 1

Unclassified
    Morgan
    Mesa
    Rio Grande
    Total Counties: 3

CONNECTICUT

High Impact
    Total Counties: 0

Unclassified
    Hartford
    Total Counties: 1

FLORIDA

High Impact
    Marion
    Hillsborough
    Manatee
    Orange
    Lake
    Polk
    Palm Beach
    Hendry
    Collier
    Dade
    Total Counties: 10

Unclassified
    Alachua
    Gadsden
    Madison
    Swanee
    Pasco
    Hardee
    Volusia
    St. Lucie
    Sumter
    Lee
    Total Counties: 10

GEORGIA

High Impact
    Colquitt
    Coffee
    Total Counties: 2

Unclassified
    Berrien
    Tift
    Brooks
    Appling
    Bulloch
    Tattnell
    Pierce
    Peach
    Total Counties: 8

IDAHO

High Impact
    Canyon
    Bingham
    Total Counties: 2

Unclassified
    Payette
    Cassia
    Jerome
    Minidoka
    Twin Falls
    Bonneville
    Total Counties: 6

MICHIGAN

High Impact
    Berrien
    Van Buren
    Allegan
    Ottawa
    Kent
    Oceana
    Total Counties: 6

Unclassified
    Manistee
    Grand Traverse
    Bay

Michigan Continued

Unclassified Continued
Saginaw
Huron
Tuscola
Sanilae
Lapeer
Lenawee
Monroe
Total Counties: 10

NEW MEXICO

High Impact
Dona Ana
Total Counties: 1

Unclassified
Total Counties: 0

NORTH CAROLINA

High Impact
Beaufort
Bertie
Craven
Martin
Pitt
Brunswick
Columbus
Cumberland
Duplin
Harnett
Lenoir
Pender
Sampson
Wayne
Alamance
Caswell
Edgecombe
Franklin
Granville
Halifax
Johnston
Nash
Person
Vance
Wake
Warren
Wilson
Bladen
Davidson
Forsyth
Guilford
Lee
Moore
Randolph
Robeson
Rockingham
Stokes
Yadkin
Ashe
Henderson
Surrey
Wilkes
Total Counties: 42

Unclassified
Hertford
Jones
Greene
Onslow
Durham
Orange
Chatham
Hoke
Richmond
Union
Alexander
Buncombe
Burke
Haywood
Iredell
Lincoln
Madison
Total Counties: 17

OHIO

High Impact
Total Counties: 0

Unclassified
Putnam
Wood
Sandusky
Seneca
Huron
Darke
Lorain
Wayne
Stark
Total Counties: 9

OREGON

High Impact
Clackamas
Linn
Marion
Multnomah
Washington
Yamhill
Hood River
Wasco
Umatilla
Jackson
Malheur
Total Counties: 11

Unclassified
Columbia
Lane
Polk
Union
Douglas
Jefferson
Klamath
Total Counties: 7

TEXAS (Unadjusted and Adjusted)

High Impact (Unadjusted)
Williamson
Hidalgo
Lamb
Lubbock
Total Counties: 4

Unclassified (Unadjusted)
Fayette
Gonzales
Lavaca
Falls
Hill
McLennon
Bexar
Frio
Duval
San Patricio
Cameron
Wharton
Dawson
Gaines
Bailey
Hockley
Terry
Floyd

Hale
Comanche
Johnson
Collin
Ellis
Smith
Van Zandt
Total Counties: 25

High Impact (Adjusted)
Hidalgo
Lamb
Lubbock
Total Counties: 3

Unclassified (Adjusted)
Fayette
Williamson
Gonzales
Lavaca
Falls
Hill
McLennon
Dimmit
Zavala
Val Verde
Maverick
Webb
Bexar
Frio
Uvalde
Bee
Duval
Nueces
San Patricio
Cameron
Starr
Willacy
Wharton
Dawson
Gaines
Martin
El Paso
Castro
Deaf Smith
Parmer
Bailey
Hockley
Lynn
Terry
Floyd
Hale
Comanche
Johnson
Collin
Ellis

Texas Continued

Unclassified (Adjusted) Continued
        Smith
        Van Zandt
        Total Counties: 42

WASHINGTON

High Impact
        Clark
        King
        Pierce
        Skagit
        Snohomish
        Whatcom
        Benton
        Chelan
        Okanogan
        Yakima
        Douglas
        Franklin
        Grant
        Walla Walla
        Total Counties: 14

Unclassified
        Clallam
        Cowlitz
        Lewis
        Klikitat
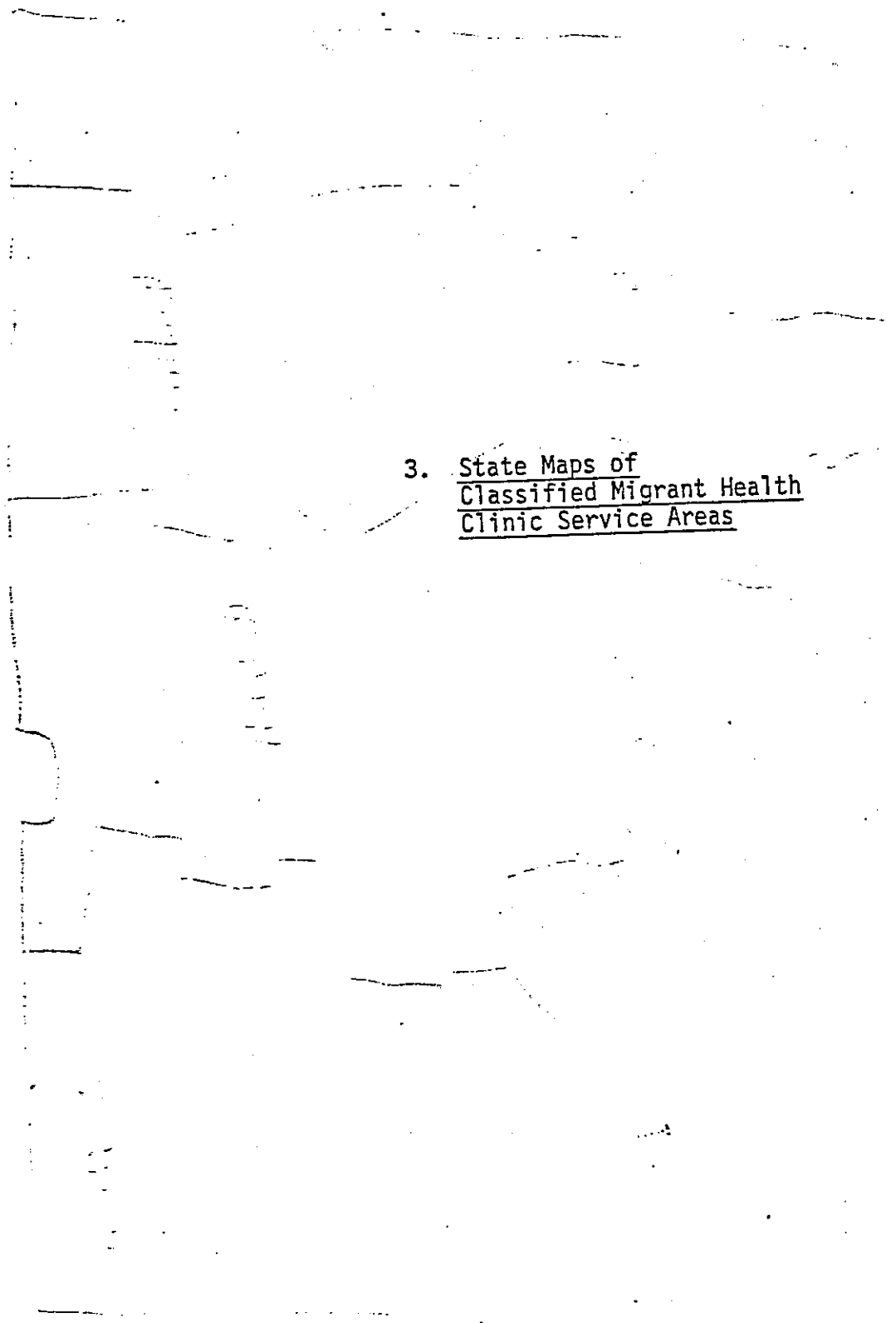   Spokane
        Stevens
        Adams
        Total Counties: 7

2. <u>State Maps of</u>
   <u>Classified Counties</u>

3. State Maps of
   Classified Migrant Health
   Clinic Service Areas

D.   <u>State Tables</u>

IV. PHASE III

CLASSIFY EACH UNCLASSIFIED COUNTY AS A
HIGH OR LOW IMPACT AREA

A. Conference of Data Experts

Plans for Phase III of the original methodology called for a meeting of data experts to be held to design a more exacting enumeration process to designate counties unclassified through Phase II as high or low impact. Such counties were ones in which ETA 223 and COA calculated figures (and for Texas, Migrant Education -- MSRTS and GOMA figures) fell between a range of 3,000 and 5,000 MSFWs at peak presence or counties in which these data sources disagreed as to high or low impact.

The idea behind the Conference was that by bringing together experts in varied disciplines with different knowledge and experience concerning demographic data development on MSFWs and other hard to count populations, the group, as a unit, could devise an appropriate methodology. The consultant selected the list of invitees to include a wide range of disciplines and a geographic diversity and, thereby, familiarity with different segments of the MSFW population. Representatives from several government agencies which deal with data development and/or information on MSFWs were also invited.

Before the Conference, participants were sent background discussion papers. These included (1) "Estimating the Target Population for the Migrant Health Program, 1983," a description of past efforts to identify the Program's target population and information on the current methodology, (2) "Problems and Methods in Counting MSFWs" and (3) List of Participants. Copies of these papers are included at the end of this section. Those invited were also encouraged to request additional background information.

The Conference was held in Rockville, Maryland on July 28-29, 1983. Twelve people attended as well as others from the OMH and from various divisions within PHS. Three additional people had been invited but could not participate.

The Conference was very productive involving lively discussion and participation by all who attended. A summary of meeting discussion is presented at the end of this section. In brief, participants were dissatisfied with the parameters they were given to design the data collection method, specifically the fact that only approximately $10,000 was available per state. The group was also concerned that the only information sought was high or low MSFW impact. They felt that much more information was necessary to properly plan any health-related service. As a united group, they concluded that a survey specifically tailored to MSFWs was the only way to assure gathering accurate and complete data, but they felt development of a survey was not possible due to the limited budget proposed for the methodology.

The groups decided that with the available funds all that could be done was to use existing data on MSFWs and reinterpret it to reflect the presence or absence of this population group. They settled on a technique which aimed specifically at defining heavy and low MSFW concentrations without providing exact or even a range of actual numbers. They suggested that information be gathered from a variety of sources and the statistical technique of discriminate analysis, or its companion logistic regression, be applied to determine a set of predictor variables which would designate high and low MSFW presence by county.

B. Description of Discriminate Analysis

As defined by David G. Kleinbaum and Lawrence L. Kupper, discriminate analysis involves "Developing a rule or discriminate function, based on the measurement obtained on each [individual] which will help to assign some new individual

to the correct population when it is not known from which of the two populations the individual comes." (Applied Regression Analysis and Other Multivariable Methods, Duxbury Press: North Scituate, Massachusetts, 1978, p 414). In regard to the methodology under development, the "individual" of the definition is a "county" and the "two populations" are "high MSFW impact area" and "low MSFW impact area." The process uses data on known high and low impact areas to pinpoint predictors of similar areas. These predictors correlate such that when they are considered together they can determine the correct population classification (high or low impact).

Predictors are variables related in some way to the populations. Variables, for this process, would be pieces of data from a variety of sources; such as COA number of agricultural workers employed under 150 days, a combination of acres planted and sales of crops, etc. This information would be in raw form and not adjusted in any way. Variables would be gathered, for each of the known high or low impact counties from many sources which relate to MSFWs. These variables would be fed into a computer and used in a variety of combinations until the best equations are found which determine high and low impact. The variables in this equation interact on each other and, therefore, may not by themselves determine high or low impact. Once a variable becomes part of the discriminating equation, it can be said to become a predictor of high or low impact. For simplicity, such variables are called "predictors" in this discussion, and rather than refer to the variable itself, it refers to the information source form which the data comes, e.g., the predictor is not the number of agricultural workers employed under 150 days but the COA itself. Determined predictors can be tested on a certain number of known high and low impact counties to see if the predictors properly classify the counties.

Any variable can be considered a potential predictor, but the final equations will include many fewer true predictors of high and low impact than the complete list of variables tried in the computer analysis. Accordingly, data will have to

be gathered from fewer sources once the predictors are determined. Gathered

data from the predictors on each unclassified impact county will be compared by

the computer to established predictors for high and low impact to determine if the

county in question is more similar to a high impact area or a low impact area.

The computer will also generate an error rate for its classification of the county.

Discriminate analysis predictors will be determined by state from known impact

status counties and will be used only with other counties in the same state. This

was specified by the Conference of Data Experts in consideration of potential

differences among states in the characteristics of MSFWs and agricultural production.

There was some discussion of grouping states into regions in some manner and developing

predictors for the counties in the region. This might be considered when only

a small number of known impact counties can be identified in a particular state

or data sources in one state are weak.

Discriminate analysis places no numerical value on the data applied to deter-

mine high and low impact. In fact, the terms "high impact" and "low impact" are

relative to the predictors and might better be defined as counties "with a lot

of MSFWs" and counties "without a lot of MSFWs." A statistical process related

to discriminate analysis, called logistic regression, may also be tried as part

of the methodology. Logistic regression is similar to discriminate analysis in its

determination of predictors, but these are used to stratify counties from high to

low impact. The gradation may include from three to six or more steps which range

from the highest impact to the lowest impact. This might provide a better

breakdown of the degree of impact, but, as with discriminate analysis, no numbers

or ranges could be attahced to the grades of impact other than to say that one

implied greater MSFW presence than the next.

Use of discriminate analysis changes the methodology to define the target

population of the Migrant Health Program somewhat in that it encompasses Phase II

and Phase III. Originally, Phase II placed counties in three categories: high impact, low impact and unclassified; and Phase III was designed to place the unclassified counties in one of the other two categories. Discriminate analysis classifies all counties as high or low impact based on predictors derived from known impact counties in the same state. This changes the methodological process to make Phase II defining known high and low impact counties, gathering data on these counties and determining predictors. Phase III then becomes application of the predictors on the other counties in the state to classify them as high or low impact.

C. Unanswered Questions About Discriminate Analysis

None of the data experts at the Conference described discriminate analysis as the ultimate methodology to estimate the Migrant Health Program target population. They stressed that it is a statistical technique which uses existing data to predict conditions (high MSFW impact). They also strongly indicated that this is not the best method to do the job (the ideal method, they insist, is a survey). However, given the methodological parameters, and particularly the funds available, discriminate analysis may be the best that can be expected.

Discriminate analysis may perform well in choosing high and low impact areas. Success of the method and the degree to which determined predictors work as expected will not be known until actual data are placed in the computer. However, it is important to realize the limitations of discriminate analysis and the questions which still remain about its use before this process is employed so that a more precise picture of its true value can be obtained. The following points remain as questions or comments on this method to define the target population for the Migrant Health Program.

1. Is it possible to use discriminate analysis on a state when only a few counties are known to be high or low impact?

2. Is it possible to use a source if it is missing data on one or more counties?

3. Is it possible to use a source if the quality of data from county to county is uncertain?

4. Will discriminate analysis clearly designate between high and low impact counties or will it simply create three categories -- definitely high impact, definitely low impact and unclassified, similar to the current methodology Phase II?

5. Will it be possible to group states into regions to determine predictors and thereby gain predictors for states with weak data sources or states with fewer number of known high or low impact counties? If this is possible, how will states be grouped -- by migrant stream, by similarity of crop types, by length of growing season, etc?

6. Will it be possible to use logistic regression on each state to allow stratification from high to low impact or must states be grouped before this statistical process can be used? How will the cutoff point between high and low impact be determined if stratification is used?

7. Can the phrases "high MSFW impact" and "low MSFW impact" be used to describe the two categories determined by discriminate analysis; i.e., what is the exact definition of the two "populations" discriminate analysis will determine?

8. Since discriminate analysis will technically not define these two categories as over or under 4,000 MSFWs at peak presence, is it reasonable to define high and low impact areas using these figures during Phase IV of the methodology: local review?

9. Is it possible to use ETA 223 and COA data adjusted to represent MSFWs at peak to determine known high and low impact counties and then reuse these data in their original form as possible predictors? If not, what other method can be used to determine known high and low impact counties?

10. Is there any way to attach numerical ranges to the two categories defined by discriminate analysis or the gradations determined by logistic regression.

D. Mechanics of Applying Discriminate Analysis

1. Personnel

The following recommendations present the views of the consultant concerning the next steps which need to be taken in the development of the methodology to

estimate the MSFW target population of the Migrant Health Program. Three agents are suggested for involvement in application of this process:

1. A contractor employed by the OMH to perform most of the data gathering and display tasks necessary and to write a manual describing the entire methodological process.

2. A statistician hired by the OMH who is highly experienced with discriminate analysis and other similar methods of data analysis (e.g. logistic regression) to advise on the technical aspects of the methodology and be responsible for computer analysis of gathered data to derive appropriate predictors of high and low impact areas.

3. A consultant hired by the OMH who is familiar with this methodological development to serve as a liaison between the other agents and the OMH, assure that tasks progress and problems are solved and guarantee that the end result is the development of a methodology which will define the Migrant Health Program target population including a written manual which can be used by others to enact this methodology.

The contractor should be familiar with basic research techniques and should be experienced in data gathering and analysis. A background including work with MSFWs or experience with gathering demographic data on hard-to-count populations would be helpful. Contractor staff should be able to follow through on investigation of data sources and have an ability to perform detailed work of an exacting nature. They should also be experienced in writing instruction manuals.

The statistician needs to be highly skilled in computer manipulation of data and experienced in similar problems; i.e., experimenting with statistical techniques to define an unknown population. He/she should be familiar with large-scale analysis and preferrably have worked with MSFWs or other hard-to-characterize populations. The statistician should be able to perform appropriate technical processes, be able to explain to non-statisticians the details of what has been accomplished and assist with the write-up of technical details for the instruction manual. He/she should be able to relate statistical data manipulation to the reality of MSFWs, in different regions of the country, with various characteristics.

The consultant should have an intensive knowledge of MSFW data and information

gathering efforts which have been attempted or might be experimentally tried with this population. He/she should be well aware of the known characteristics of MSFWs, migrant travel patterns and the regional and state differences of the individuals who perform farmwork. The consultant should be able to design, direct and guide a systematic work plan as well as identify trouble spots and propose solutions. He/she should be familiar with past efforts of the OMH to enumerate its program target population and have a working knowledge of the present methodology development effort. The consultant should be able to write technical details at a level of understanding appropriate to someone unfamiliar with the subject.

2. Application of the Methodology

Five steps should be used to complete this methodological development: applying discriminate analysis or other statistical techniques on gathered data to determine the high or low MSFW impact status of counties, instituting a local review process to validate county classifications (Phase IV of the original methodology), and developing an instruction manual on how to perform the entire methodology. These steps are described below. The division of tasks among the contractor, statistician and consultant are identified. It is suggested that the ten states used thus far in this methodological development continue as samples for the remainder of the process. It had been proposed that this sample be reduced to five states; however, it is felt that there is sufficient time and funding to continue to test all ten, plus a need to sample as many states as possible to assure the methodology is accurate and effective.

Step 1: <u>Pinpoint high and low MSFW impact areas and appropriate data sources and collect identified data.</u>

The consultant and statistician should confer on the number of known high and low impact counties which must be identified in each state to assure an accep-

table sample for development of predictors. Also, the appropriate number of known high and low impact counties to be set aside for a test of the predictors should be specified. Decisions regarding these points should be finalized through discussion with the OMH. Using Phase II data on the sample states, the consultant should identify the appropriate number of high and low impact counties.

The consultant will specify a list of primary data sources (variables) for each of the states and will identify the form in which the data should be collected. This task will be performed through consultation with the statistician. Such sources may include: ETA 223; COA; crop sales and acreage planted in fruits, vegetable, tobacco and nursery plants; and 1980 Census of Population information. The consultant will also prepare a list of secondary sources from which variables should be collected if the information is available, complete and found to be relevant. These may include: vital statistics, migrant health clinic MSFW patient family member numbers, MSFW enrollment in Women Infant and Children (WIC) programs, figures for migrant families receiving Food Stamps, statistics from the Immigration and Naturalizations Service, Unemployment Insurance claims paid for farm laborers, Social Security records of MSFWs, and MSRTS enrollment.

The contractor will, by telephone, contact agents in each of the states to collect primary variables for every county and to judge whether secondary sources of information are appropriate for use. At the same time, the contractor will explore other data sources such as special surveys, state or local data collection systems, studies by educational institutions, etc. Each source should supply data by county and should relate to MSFWs in some way. It is not necessary to secure variables form the same sources in each state as the states will be analyzed separately to determine appropriate high/low impact predictors. The contractor will confer with the consultant on each new source discovered before expending effort to collect actual pieces of information.

At this step in the methodology, it is really only necessary to collect variables on the identified known high and low impact counties. This information will be used to determine relevant predictors and only data from these predictor sources will need to be collected to classify other counties as high or low impact. However, for many sources, the contractor may find it just as easy to ask for data on every county. All information should be collected by telephone and by mail without necessitating a site visit.

The contractor will work with the consultant to decide on an appropriate display for variables for each identified high and low impact county. At this point, it would not be productive to display all of the data for the unclassified counties as not all of the sources will be determined valid predictors. For some sources, it may be easier to obtain a data tape which can be programmed into the computer during analysis. Under the guidance of the consultant, the contractor will document in detail the weaknesses of each source. This includes specification of missing information (e.g., some counties have no data), problems discovered in the methodology used by a source or in actual field collection of information, differences between the definition used by the source and the one used in the Migrant Health Program, and other problems. If the contractor feels a source provides particularly accurate data, this should also be documented.

The final products of this Step should be: (1) a list of known high and low impact counties in each state with some set aside for later testing, (2) a list of primary and secondary variables for each state noting how information was obtained from each source and any weaknesses or strengths in the data, and (3) for at least the identified known high and low impact counties, an array of variables by county.

Step 2: <u>Computer analyze gathered data using various statistical techniques to obtain predictors of high and low impact areas and test these predictors.</u>

The statistician will work through the Parklawn Computer Center to place the arrayed county data in the computer and use various combinations to derive a set of predictors tied to specific data sources which will designate high and low impact counties. The statistician will apply discriminate analysis, logistic regression or other techniques to the data to establish predictors for high/low impact or develop a stratification from high to low impact. First run analysis will be by state, with later analysis grouping states with similar characteristics to attempt to identify regional predictors. It is expected that the statistician will try every possible combination of variables to identify the best predictors.

In performance of this analysis, the statistician will work with the consultant to explain and confer on procedures and results. The OMH will also be involved in the process.

Once high and low impact predictors are determined by state, they will be tested on counties known to be high or low impact but which were not used in determination of the predictors. Success in identifying the proper impact area of test counties will place greater confidence in the ability of the predictors to discriminate between high and low impact. The statistical program, itself, also provides a level of confidence for the predictors.

In making the final determination on predictors, the statistician and consultant will examine what has been written concerning the weaknesses and strengths of the data sources to which the predictors belong. The final product of this step will be a list of predictor data sources for each of the ten states which should determine high and low impact counties and the degree of confidence attached to such predictions.

Step 3: <u>Collect additional data, if necessary, and use established predictors to analyze information to designate unclassified counties as high or low impact.</u>

When relevant predictors have been determined, the number of sources from which data must be gahtered will be greatly reduced. For example, if information from all the primary and secondary sources had been gathered this would result in 12 variables. Only four of these sources, however, may be found to be true predictors of high and low impact. It was necessary to gather information from the other eight because it was impossible to know which would perform best before computer analysis.

The contractor will collect information on the unclassified counties in each state from the appropriate predictor sources. In some instances, the contractor may already have collected this information during Step 1. Similar to earlier data collection efforts, the contractor will document any weaknesses or particular strengths found in the data sources. The consultant will oversee this process, solving any problems which arise, and conferring on the final data display.

The statistician will take the arrayed data on unclassified counties and analyze it in light of the determined predictors for high and low impact. He/she will note and document any problems which arise during this process. The final result of this step will be a preliminary listing for each of the ten states of high and low impact counties and/or if it was possible to stratify from high to low impact, counties will be classified accordingly.

Step 4: <u>Perform a local review of counties classified as high or low impact and adjust county classifications if necessary.</u>

The local review process is described in detail in the Final Report section "Phase IV, Verify the Classification of Counties as High or Low Impact Through Review by Local Sources." The tasks described should be performed by the contractor in consultation with the consultant. Returned Local Review Forms should be independently examined by the contractor, the consultant, and the statistician, as described in the write-up. All three parties should be involved in any decision

to change a county classification as a result of local review information. All activity in regard to action taken from local review comments should be thoroughly documented. The result of this step will be a final listing of high and low impact counties in each of the ten states.

Step 5: <u>Write an instruction manual on how to apply the developed methodology to counties in the rest of the country</u>.

The contractor, consultant, statistician and OMH will all confer on the information needed in the methodology instruction manual and the audience for whom it should be written. The contractor will be the primary writer of this manual with guidance from the other agents. The contractor will refer to the Phases I and II write-up of this Final Report in describing how to determine known high and low impact counties from which to derive predictors. A more simplified "how-to" version of the information presented should be written.

The contractor will describe tasks necessary to identify and collect information from primary data sources. The contractor will seek guidance from the statistician in writing the technical computer analysis steps used to determine approrpiate predictors, the test of these predictors and the application of them to data from unclassified counties. The contractor, in discussion with the consultant and statistician, will describe data deficiencies found to be serious enough to disqualify a source from consideration as a predictor and characteristics of sources which might make them good predictors. (This will be determined from the documentation of data weaknesses and strengths found in the types of sources used from the sample states.)

The contractor will confer with the consultant to write a final version of the local review process which considers the problems encountered with the sample states and establishes a systemized method for reviewing and processing local comments. Examples of the type of data presented by local reviewers should be available based on experiences with the ten sample states.

The draft instruction manual should be reviewed by the consultant, the statistician and the OMH for accuracy, completeness, clarity and simplicity. Needed changes should be made and then its usefulness tested on a few members of the audience for whom it is intended. Reviewers should not have been involved with the development of this methodology in any way, and their comments should be incorporated into manual revisions. A good test of the manual's value might be to have the reviewers explain the methodological process to the OMH or the consultant based on the reviewers understanding from the manual.

Papers From The

Conference of Data Experts

ESTIMATING THE TARGET POPULATION

FOR THE MIGRANT HEALTH PROGRAM, 1983

The Migrant Health Program is designed to provide primary health care services to migrant and seasonal farmworkers (MSFWs) and their dependents, with particular attention to assisting migrants as they travel from their home base through the migrant stream performing seasonal farm work. Most services are provided through migrant health centers and clinics located throughout the country in rural agricultural producing areas. Local programs are administered by a private, non-profit organization or by state or local health departments. The size and sophistication of individual programs varies considerably with some offering a full range of comprehensive care through an extensive on-site medical team and others contracting for services through local physicians. A few programs are seasonal, providing health care services only during the migrant influx. Most operate year-around serving seasonal farmworkers and home-based migrants during the off-agricultural season. Many programs have other types of funding which allow them to provide health care services to local population segments besides MSFWs, particularly low income individuals.

The Migrant Health Program uses the following definition of a MSFW:

> The term "migratory agricultural worker" means an individual whose principal employment is in agriculture on a seasonal basis, who has been so employed within the last twenty-four months, and who establishes for the purposes of such employment a temporary abode.
> The term "seasonal agricultural worker" means an individual whose principal employment is in agriculture on a seasonal basis and who is not a migratory agricultural worker.
> The term "agriculture" means farming in all its branches, including --
> (a) cultivation and tillage of the soil,
> (b) the production, cultivation, growing, and harvesting of any commodity grown on, in, or as an adjunct to or part of a commodity grown in or on, the land, and
> (c) any practice (including preparation and processing for market and delivery to storage or to market or to carriers for transportation to market) performed by a farmer or on a farm incident to or in conjunction with an activity described in subparagraph (b).

The Migrant Health Program is legislatively mandated to periodically re-estimate the target population eligible for services. The information is used to justify the existence of the Program as well as determine whether centers and clinics are appropriately placed (in terms of number to be served), and judge if more local programs are needed. Target population estimates may also be used to allocate Migrant Health Program funds to individual centers and clinics based on the need for services.

In 1973 and again in 1978, the Office of Migrant Health produced a document titled Migrant Health Program Target Population Estimates which provided the enumeration information which was needed. Authors of both documents stressed the extreme difficulty of deriving reliable population estimates by county, as were required by the Office of Migrant Health. Lacking funds for primary research, the 1973 Report used existing secondary source information through which to obtain estimates. Monthly estimates produced by state departments of employment for number of seasonal (local) and migrant (non-local) workers employed in agriculture in specific geographic areas were the principal resource for these numbers. Figures were adjusted for an error rate and factors were applied to enlarge the estimates to include dependents. The error rate used was constant throughout the data but the dependent factors varied regionally or by state and were based on available independent studies located by the researchers. Dependent factors differed for migrants and for seasonals.

The 1978 Target Population Report essentially used the same data sources and methodology including no change in the earlier dependent factors. The major methodological differences were a re-estimate of the data error and the use of any other locally produced data, particularly survey information, to suppliment the final estimates.

Both Reports provided estimates of peak number of migrants, peak number of seasonals and peak number of MSFWs present in each county in the United States determined to contain a significant number of seasonal agricultural workers. A local review process was used to catch gross data errors before finalization of the estimates; however, many individuals, including the authors of both Reports and the Office of Migrant Health, were skeptical of the accuracy of the final county estimates due to the method used.

In 1982, the Office of Migrant Health began to again consider the problem of estimating MSFWs, given a precedent of re-estimating the Migrant Program target population every five years. Migrant Health was determined to design a new way to derive the data needed, due to dissatisfaction with the methodology of past Reports and the feeling that the different systems used by individual state departments of employment to estimate monthly MSFW employment figures had become less scientific assuring greater inaccuracy in estimates based on these numbers. Discussion with a consultant hired to assist in the design, Dr. Alice Larson, and with knowledgeable individuals at the Bureau of the Census and U.S. Department of Agriculture, helped refocus the estimation effort in regard to the specific information needed. This resulted in data collection guidelines which differ slightly from the past two Reports.

It was determined that the Office of Migrant Health's real need was to identify "high impact areas," which are mandated to be served first with Migrant Health Program funds. A high impact area is: "a health service area or other area which has not less than four thousand migratory agricultural workers and seasonal agricultural workers [and members of their families] residing within its boundaries for more than two months in any calendar year." These areas then became the objective of MSFW enumeration.

Despite the high impact area definition requiring two peak periods, it was decided to obtain MSFW estimates at maximum presence, as it is difficult to determine dual impact times and past Reports provided only one peak number. Information,

as in past Reports, was to be gathered by county; however, the Office of Migrant Health plans to group counties to form logical health service areas. The primary differences between the proposed effort and past efforts are (1) an objective to identify counties as high impact (4000 or more MSFWs at peak) or low impact areas, not an exact number, and (2) no effort to break the population into separate groups of migrants and seasonals unless this has a bearing on determining high impact counties.

The current enumeration effort involves: a period of methodological development, testing and evaluation of the proposed method, and use of the final system throughout the country. The process combines existing data sources with a more exacting enumeration. Ten sample states have been chosen for the developmental stage including both in-stream and home base areas in the three classic migrant stream patterns; east coast, west coast and mid-continent. The states are Colorado, Connecticut, Florida, Georgia, Michigan, New Mexico, North Carolina, Ohio, Texas and Washington. The methodology being developed consists of four phases, as described below. At this point, work has been completed on Phases I and II.

Phase I

Purpose: Identify counties which employ seasonal farm labor.

The only sources of existing data which make estimates of the MSFW population
by county on a systematic basis were identified and used to list those that
employ seasonal agricultural labor. These include: ETA Form 223 monthly
estimates of employed migrant and seasonal workers generated by each state
department of employment as required by the U.S. Department of Labor (the
primary source used in the 1973 and 1978 Target Population Reports); Census of
Agriculture yearly estimates of agricultural laborers working under 150 days
on individual farms, developed as sample survey data by the Bureau of the Census;
and migrant student enrollment numbers within Title I Migrant Education Programs
in each state as generated from the Migrant Student Record Transfer System
(MSRTS), a computerized student record system headquartered in Little Rock,
Arkansas.

None of these data sources use the same definition nor the definition of MSFWs
used by the Migrant Health Program. One source provides estimates on a monthly
basis, whereby peak numbers can be easily identified. The other two sources
give year-around numbers with no indication of percent present at peak. Two
sources provide data on employed workers only, without consideration of dependent
non-workers. Children enrolled in Migrant Education Programs are the target
of the third source which does not consider related parents, children too young
to be enrolled in the Program and children not enrolled in the program.

ETA 223 data used in this Phase were from 1980-1981 depending upon the state.
Migrant Education numbers covered 1981, and Census of Agriculture estimates were
for 1978.

Phase II

Purpose: Divide counties in which seasonal agricultural workers are employed
into one of three categories: high impact (4000 MSFWs or more at peak), low
impact (under 4000 MSFWs), and uncertain impact.

Using various factors, each of the source figures were converted to peak number
of MSFWs, including dependents, and the results compared to place each county
in one of the three impact categories. Review of MSRTS data revealed apparent
discrepancies between the completeness of records from county to county. It
was found that although students enter into the system when they enroll in school,
many others are brought in through the efforts of local recruiters. Therefore,
the accuracy of the numbers depends upon the aggressiveness of this local agent.
In addition, MSRTS figures do not exist in some areas which lack a Migrant
Education Program, and it is necessary to apply many factors to these data to
convert them to peak MSFW presence, including: percent of children not in the
system, number of children in the system per family, adults per family, and
number of children too young to be in the system. For all of these reasons,
MSRTS figures were dropped as a source for data comparison.

Factors to determine the number of dependents per worker represented in the two
remaining sources were pulled from the 1973 Migrant Health Program Target
Population Estimates Report. Additionally, a data error rate was applied to
ETA 223 numbers and two percentage factors were used on Census of Agriculture

figures to convert them to peak employment from year-around tallies. For the latter source, the application of separate peak percentages created two estimates. These plus the converted ETA data provided three MSFW peak numbers for each county. These estimates were then compared to determine whether each county could be classified as a high impact, low impact or uncertain impact area.

Phase III

Purpose: Determine whether each area identified as uncertain impact should be classified as a high or a low impact area.

In July, 1983, a group of experts from throughout the country will determine the method to be used in a more exacting enumeration of the uncertain impact areas to classify these remaining counties as high impact or low impact. Experts will include demographers, statisticians and agricultural economists, as well as those familiar with the separate migrant streams and with the primary sources of data on MSFWs.

This group will consider the following enumeration schemes and others in its determination of the most feasible, cost effective and timely method: sample scientific or non-random survey, conversion of need for hand labor -- as dictated by crops grown and acreage planted -- into number of workers and dependents, adaptation of a capture-recapture methodology, and use of local knowledgeable sources to determine proper placement of counties as high or low impact. The methodology developed by the group will be refined and tested on the uncertain-impact areas in each of the sample states. This new input will then place such areas in either the high or low impact category.

Phase IV

Purpose: Verify the categorization of counties as high or low impact through review by local sources.

The classification of counties in each state as high or low impact will be examined for accuracy by the Regional Migrant Health Program staff, by migrant health center and clinic personnel, and by other local agents. For any county felt to be misplaced, the reviewer will be asked to present evidence as to the proper classification or documentation that the sources used to derive the estimate were inaccurate to the point of causing miscategorization.

PROBLEMS AND METHODS IN COUNTING MSFWs

I. MSFW Counting Problems.

A. Definition: There is no one definition of who is/is not a "migrant" or "seasonal" farmworker in common use by service providers, policy makers, those who enforce regulations and data gatherers. This means the target population cannot be defined, and a data source prepared for one agent may not be relevant to another.

B. Movement: MSFWs move from job to job and migrants move across the country. They are difficult to find and interview at work. Migrants have no permanent residence while working and so are hard to find at a living site.

C. Duplication: Because MSFWs constantly move, it is easy to count one person more than once.

D. Population Factors: Total numbers cannot easily be determined from information on population segments. For example: the total population cannot be estimated from the known number working on specific days because of duplication of counts. Family size cannot be projected from number of children in school or number of family workers in the field because children in school and number of family workers varies depending on community resources, family economic situation and other factors. The total population cannot be estimated from the number who seek social services because it is unknown how many do not seek assistance.

Sampling Frame: There are no complete listings available of work or residence sites for MSFWs. It is often impossible to make such exhaustive listings. For example: there are no complete labor camp lists or total agreement over the definition of a "labor camp." Seasonals do not necessarily live in one neighborhood. In many areas, there is little housing for migrants who are forced to live wherever they can including in their cars. Employers of MSFWs do not have to be licensed and many do not hire MSFWs for a sufficient length of time to be required to report their social security earnings.

F. Change: There is a feeling that the MSFW population is constantly changing depending on the economic situation of the country and labor needs. MSFWs are not like veterans -- once a person is a migrant or seasonal that person is not always a migrant or seasonal.

II. Possible Methods to Count MSFWs.

A. Demand for Labor: Examine the number of seasonal workers required for agricultural production, apply dependent factors and thereby estimate the number of MSFWs based on labor needs.

1. Problems:

- counts including several crops duplicate workers.

- different production capability of types of workers and varieties of each crop can affect labor demand calculations; e.g., picker experience may affect work speed, height and spacing of trees may affect production time.

- the factor(s) for dependents per worker (household size) is unknown.

- labor demand calculations means home-based migrants who do not perform seasonal farm work at home are not counted.

- MSFW definition problem.

2. Benefits:

    = information to determine labor needs -- crops, acreage, estimate of average person-hours to produce crops -- is readily available with reasonable accuracy.

    peak MSFW presence can be estimated given the normal heavy periods of labor needs per crop and the usual time when various crops would be worked.

    - yearly updates of MSFW number estimates might be possible based on changes in crops, acreage and person-hours needed for production (also given a scheme to determine change in dependent factors).

B. Survey (scientific sample or non-random): Select contact sites and interview MSFW respondents, relative to a sampling plan which can estimate the total population based on the information gathered.

    1. Problems:

        - identifying sampling frames.

        - achieving access to survey sites and respondents.

        - knowing accurate sample size or a reasonable number and variety of non-random sites.

        - count duplication if the survey extends for a period of time.

        - constant turnover of respondents if the survey extends for a period of time.

        - must obtain an extensive staff to conduct the survey or train local personnel.

        - field control to guarantee uniformity of survey is a necessity.

        - this method is costly.

    2. Benefits:

        - may have a high degree of accuracy if the survey is performed well.

        - provides the opportunity to collect other demographic data besides numbers.

- no definition problem as this is controlled through survey screening questions.

C. Capture-Recapture: Capture, tag and recapture respondents through a series of possible contacts, then estimate the population using log linear models best fitted to the data obtained.

 1. Problems:

 - same as "B" Survey Problems.

 - need a capture-recapture "tag."

 - the method assumes that the observed population is the same as the unobserved.

 - some individuals may avoid being captured or recaptured thereby skewing the population which is identified.

 2. Benefits:

 - might be quicker than a survey.

 - the method has the potential to provide a more accurate numerical estimate than a survey.

 - as with a survey, original capture of respondents could gather additional demographic data.

 - no definition problem.

D. Count of MSFWs: Identify a sample of representative sites and actually count MSFWs.

 1. Problems:

 - hard to assure all who are counted are MSFWs according to a prescribed definition.

 - difficult to select representative sites.

 - the factor(s) for dependents per MSFW counted (household size) is unknown.

 - count duplication if the count extends for a period of time.

 2. Benefits:

 - possibly easier access to sites than for a survey.

 - if problems can be solved, method concentrates only on numbers and would be accurate, very direct and involve primary research.

E. Expert Opinion: Survey local experts knowledgeable of MSFWs as to their best estimate of the population.

 1. Problems:

- not all "experts" have the same reference frame -- see the same segments of the population.

- definition problem.

- estimating numbers or ranges is difficult for many people.

- may be least accurate method as count is based on soft information.

2. Benefits:

- do not need exact MSFW number, but just high or low impact counties, as called for in the Migrant Health methodology.

- this system is quick rnd cheap -- can probably be done by telephone or mail.

- "types" of experts are easily identifiable and would probably be cooperative and accessible.

AGENDA
July 28-29, 1983
Conference Room M
Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20857
301-443-1153

7/28	9:00 a.m.	Opening Remarks and Welcome Dr. Michael E. Samuels Director, Migrant Health Program
	9:30 a.m.	General Introduction Enumeration of Migrant and Seasonal Farmworkers (MSFW) - Background
	10:30 a.m.	Break
	10:45 a.m.	Discussion of participants' experiences with hard to count populations
	12:00 Noon	Lunch
	1:00 p.m.	Discussion of Methodological Plan Development
	5:00 p.m.	Adjournment
7/29	9:00 a.m.	Continuation of discussion of methodological plan development
	10:30 a.m.	Break
	10:45 a.m.	Continued Discussion
	12:00 Noon	Lunch
	1:00 p.m.	Finalize methodological plan for enumeration of MSFW population
	3:00 p.m.	Participants' critique of Phase I & II of MSFW methodology
	4:30 p.m.	Closing remarks
	5:00 p.m.	Adjournment

MSFW DATA CONFERENCE
JULY 28-29, 1983, ROCKVILLE, MARYLAND
MEETING SUMMARY

The conference to design a system to classify specific counties in five sample
states as high or low impact was attended by twelve people (see attached list).
They represented a variety of disciplines, background and familiarity with MSFWs
and data gathering problems. All participants but one stayed throughout the entire
conference. Three other people were invited but did not attend.

The meeting began with an explanation of the four phase methodology being
employed by the Office of Migrant Health (OMH) to classify all U.S. counties as
high or low impact. (See earlier background paper.) Discussion was held concerning
the definition of migrants and seasonals employed by the OMH, description of a
high impact area, and purpose of the current target population data gathering
effort. Parameters for design of the county classification system (the task of
this meeting) were laid out as indicated below:

1. Develop a system better than the one used in the 1973 and 1978 Migrant
   Health Program Target Population reports.

2. The system is concerned with both migrants and seasonals.

3. Whenever possible, use the Migrant Health Program definition of MSFW
   (excludes farm owners/operators and their family members, loggers,
   fishermen, sheepherders, agricultural produce transporters, dairy-cattle-
   poultry workers).

4. The system should be uniform across states and migrant streams.

5. The only concern is high and low impact counties (numbers and demographic
   information would be a nice extra but are not essential to the methodology).

6. The system should not be too complex -- it may be performed by Migrant
   Health Clinic staff.

7. Available funding for the system is $10,000 per state.

8. Time to perform this system is six months.

9. The number of unclassified counties in the five states (in which the
   system being designed will be tested) are:

   - Florida: 17
   - Michigan: 8
   - North Carolina: 19
   - Texas: 47
   - Washington: 7

The group found the specification of funding available for the system
particularly troublesome. Some members immediately differentiated between an
enumeration system and a high and low impact classification system, preferring
the former. Throughout the entire meeting participants expressed a strong desire

to gather additional demographic information which they felt was essential to any program planning process. However, due to the specified monetary limit for the system, they felt limited in their ability to design a thorough data gathering effort. At one point in the meeting, the participants attempted to determine the cost of an appropriate methodology which could provide the information they felt to be essential. The concensus seemed to be that a complete census was not necessary, but a sample survey would be required. All seemed to feel confident that the Bureau of the Census could do an adequate job, but the cost would be very high. Several group members felt they could not adequately predict the cost of a MSFW survey at this time.

One representative from California indicated a survey of MSFWs to gather extensive work-related information was to begin soon using the resources of the State Employment Department for interviewers and the University of California at Davis for computer analysis. California law guarantees access to field workers during specified rest periods and a list of growers is readily available, thus cutting the cost of development of a survey frame and respondent accessibility. The entire state survey was estimated to cost $70,000, not including costs for those items mentioned above. The USDA representative, on the other hand, described a recent proposal submitted by her agency to enumerate MSFWs in 25 states, gaining information which cannot be broken down further than the state level. The cost of this effort was estimated to be several million dollars.

Participants agreed that a county as the unit of analysis is appropriate, particularly when it was explained that in actual use of the impact information the OMH planned to group counties into logical medical service areas for Migrant Health Centers. Troublesome data gathering problems discussed by the group included:

1. Applying appropriate dependent factors to a source to represent all MSFWs; e.g., non-working household members per worker; household members per child enrolled in a school program.

2. Converting year-around figures on MSFWs to peak presence.

3. Determining the percent of the population missing from a data source; e.g., number of MSFWs not served by Migrant Health Centers.

4. Locating accurate sources of data to measure non-working migrants who live in and migrate from an area but do not work there.

5. Considering all of the regional and state differences among MSFWs, yet assuring a uniform county classification system across the United States.

Meeting participants discussed several possible schemes for enumerating MSFWs as a way of classifying counties as high or low impact. These included multiple recapture, demand for labor, and survey of local experts. The first day concluded with the group exploring some type of screening process whereby a county would be determined high impact if any of several data sources indicated over 4000 MSFWs present at peak.

The second day's discussion began with the idea of comparing existing and reliable MSFW county censuses to available data sources to assure that those counties determined from these sources to be high impact agreed with the census conclusions. In this way the relationship of existing data sources, which specify a county as high or low, to the county's actual impact designation, based on the

census, could be determined. The main problem was thought to be locating enough reliable special censuses already performed on a representative group of counties throughout the country.

From this base, the statisticians in the group began to push for a computer generated solution to county classification as high or low impact. They suggested use of discriminate analysis as a tool to profile counties known to be high or low impact, and then applying determined impact predictors to unclassified counties to decide their status. Other members agreed with this process, and the group saw it as a way to dispense with many of the data gathering problems mentioned earlier. Specifically, discriminate analysis would allow data to be entered in its original form without applying dependent factors, converting year-around figures to peak, or concern for population segments missing from the data source. Each state could be analyzed separately to determine predictors, therefore, making use of exclusive information sources and considering characteristics peculiar to the state.

Sources of data suggested by meeting participants included:

1. Information used in Phase II (ETA 223 and Census of Agriculture).

2. Acres planted and sales of fruit, vegetables, tobacco and nursery products (agricultural commodities most likely to employ MSFWs).

3. Migrant Student Record Transfer System of the Migrant Education Program.

4. MSFW patient records in Migrant Health Clinics -- count would be for all family members and divided by counties where patients reside.

5. Food Stamp records for migrant enrollees and their family members -- each state welfare agency is required to keep these records.

6. Unemployment Insurance records for farmworkers receiving benefits.

7. MSFW enrollees (women and children) in local WIC programs.

8. Different pieces of information from the 1980 Census of Population.

9. Different pieces of information from state vital statistics records.

10. Other special one state data sources.

Some problems with application of the method were discussed including assurance that counties used to determine relevant predictors are definitely high or low impact, data sources which miss individual counties, sources with more reliable information on some counties than on others, and locating enough sources of data in some states. Participants explained that the technique of discriminant analysis can use any piece of information as a predictor and sometimes it is unclear why individuals predictors work, although statistically they perform the job. A few others in the group, however, worried about the use of false predictors to determine presence of MSFWs; e.g., 1980 Census data and vital statistics which only peripherally include migrants.

One caution with use of this system was presented. It would not provide any population enumeration, and in fact, would not divide counties on a high impact

number of 4000 MSFWs. Predictors determined by discriminate analysis would indicate which counties have "a lot of" MSFWs and which had "few," and could put gradations on these phrases. The OMH, in consultation with experts, would then have to determine the cut-off point for high impact.

The group concluded its discussion of this data method by specifying that the data must be gathered, arrayed, and computer manipulated before anyone could guarantee that discriminate analysis would work as a county impact classification system. Nevertheless, given the monetary resources, discriminate analysis may be the best available method. The data gathering method of first choice for the group remained a survey which could provide enumeration and other demographic data. The group wanted to consider discriminate analysis as a first step to a more extensive MSFW data gathering effort.

The last concern expressed was for the mechanics of the system which was proposed. The group felt the best scheme would include the following steps:

1. Determine assured high and low impact counties in the sample states.

2. Gather and array data on these counties, using a variety of sources. This task should be performed by an OMH contractor.

3. Manipulate data on 90% of these classified counties to determine appropriate predictors. The PHS Computer Center resources should be used for this process in close consultation with an outside statistician familiar with the method, the problem and the goals of the county classification system.

4. Determine appropriate predictors and test the application of these on the set-aside 10% of classified counties.

5. Tested and proven predictors should then be used to designate as high or low impact unclassified counties in the sample states. As in step "2," an OMH contractor should gather relevant data from these counties.

Throughout this process, meeting participants felt it would be essential to continue the current technical advisory system, using the services of a consultant who has been involved with the OMH in this data development methodology in the role as liaison between the contractor hired by OMH to gather county data, the consultant statistician, meeting participants and the OMH. This will assure continuity in the process and focus on the ultimate goal: development of a method to classify counties as high or low impact.

MSFW DATA CONFERENCE
JULY 28-29, 1983, ROCKVILLE, MARYLAND
MEETING PARTICIPANTS

1.. Steve Dann, Office of Migrant Health, Public Health Service, U.S. Department
    of Health and Human Services, Rockville, Maryland.

2. Stephen Fienberg, Professor, Department of Statistics, Carnegie-Mellon University,
    Pittsburgh, Pennsylvania.

3. Ruth Ann Killion, Supervisory Mathematical Statistician, Statistical Methods
    Division, Bureau of the Census, Washington, D.C.

4. Gontran Lamberty, Director, Research Division, Maternal and Child Health, U.S.
    Department of Health and Human Services, Rockville, Maryland.

5. Alice Larson, Private Consultant, Seattle, Washington.

6. Phillip Martin, Professor, Department of Agricultural Economics, University
    of California at Davis, Davis, California.

7. Gloria Mattera, Director, BOCES Geneseo Migrant Center, Geneseo, New York.

8. Michael Meyer, Assistant Professor, Department of Statistics, University of
    Wisconsin, Madison, Wisconsin.

9. Nagesh Revankar, Professor, Department of Economics, SUNY, Buffalo, New York.

10. Alan Ross, Professor, Department of Biostatistics, John Hopkins University,
    Baltimore, Maryland.

11. William Schlinger, Organizational Psychologist, Center for the Study of Social
    Behavior, Research Triangle Institute, Research Triangle Park, North Carolina.

12. Leslie Whitener, Rural Labor Market Section, Economic Development Division,
    Economic Research Service, U.S. Department of Agriculture, Washington, D.C.

V. PHASE IV

LOCAL REVIEW OF CLASSIFIED COUNTIES

Phase IV of the original four-phased methodology to develop 1983 target population estimates for the Migrant Health Program was designed to incorporate local review of counties in each state which have been classified as high or low impact. This Phase of the methodology should be maintained no matter what process is used to classify counties. If some data manipulation process, such as discriminant analysis is used rather than a direct survey, local review of the preliminary data will help identify any place the method may have misclassified because of errors in the local information used in the process. The system itself may generally be accurate, but it is possible that local data quirks may throw the results off within some individual counties. For example, as discussed in the Phase II Report, reliance on ETA 223 and COA employment-based data on MSFWs in Texas would have resulted in numerous county misclassifications. Local review of this information would have caught such errors and, in fact, was used to identify the causes of these errors and suggest solutions. In addition, local review will allow local agents to submit new quantitative information and comment before county classifications are considered final. This process will, hopefully, head off criticism by some local personnel who may feel they were not consulted during the classification process or who may discount all county classifications as not founded on a sound methodology due to a few uncaught classification errors.

Local review must involve a systematized process which allows local input of reliable information but screens out qualitative opinions. One problem with the 1978 Migrant Health Program Target Population Estimates report was the local review of preliminary estimates. It was conducted late in the process, was not well organized and, in several instances, seemed to consider as valid almost any challenge to the derived numbers.

From this past experience and in consideration of the value a local review phase can play in validating manipulated data, it is important to make this process

an integral part of the overall methodology and plan for its enactment. It is also necessary to predesign a system to evaluate the local input which will be received. The following discussion presents ideas for this methodological phase in some detail.

A. Basic Process

Local review of counties classified as high or low MSFW impact should be conducted by mail with telephone follow-up if necessary. Local reviewers should be sent a standardized form with instructions for its use and an explanatory cover letter. The form should ask for reviewer verification of each county class-ification and their extensive comments in regard to misclassifications. Comments should reference supportive data or such information should be sought in follow-up telephone calls. County misclassifications indicated by several local reviewers should be carefully examined and counties should be reclassified if substantial new information is presented to warrant such action.

B. Form Development

The Local Review Form sent to each reviewer should be clear and simple requiring minimal time to complete. An example of such a form and its accompanying instructions for completion are included at the end of the "Phase IV" discussion.

The form should begin by asking the name, organization, address and telephone of the reviewer. Its content should be divided into two parts. The first part titled "County Classifications," should identify the state and list counties. Each county should be marked to indicate whether it has been classified as high or low impact. Next to this designation space should be provided for the reviewer to check his/her agreement with the classification, disagreement or indicate uncertainty on the proper classification. The second part of the form, titled "Reasons for

Classification Disagreement" should have space for the reviewer to write the name of every county he/she feels is misclassified and evidence in support of their disagreement. The need to provide quantitative data on county misclassification should be stressed on this part of the form. Space should be provided for the reviewer to discuss at least ten counties. The form should conclude with the name and address of where it should be returned.

Detailed instructions on completion of the form should be sent to reviewers. These should include a one sentence summary of the purpose of the information (to classify counties as high or low MSFW impact), a definition of a high and low MSFW impact area, description of the information asked, and a date for return of the form. Within these instructions three points should be stressed: (1) the designation is for counties, not groups of counties, (however, indicate that OMH plans to group counties into MSFW medical service areas when using this information for Migrant Health Clinics), (2) county classification is for both migrant farmworkers and seasonal farmworkers plus their household members, and (3) it is crucial to present quantitative data when discussing the misclassification of a county.

A cover letter accompanying this form and its instructions for use should discuss the following items: the sponsorship of this request (OMH), the purpose of this review and how it fits into the methodological design and development of county MSFW impact data, what is asked from the reviewer and why his/her input is important, and the need for the reviewer to supply quantitative/factual information when discussing misclassification of a county to enable researchers to compare this new information to that which was used to classify the county. A return date should be given for the form in this cover letter as well as in the instructions for completion of the form. The letter should also indicate that the reviewer can request a more detailed explanation of the methodology used to preliminarily classify counties. An explanatory paper should be written for this purpose,

based on the details in this Final Report, a description of the discriminate analysis process employed, and the data sources used.

C. Form Distribution

The Local Review Form should be sent to at least the following types of individuals in every state:

* Migrant Regional Program Consultant at the federal regional office in charge of the state.

* Directors of Migrant Health Centers and Clinics in the state.

* The primary office of the state cooperative extension service (normally attached to a state university or college).

* The state employment department division responsible for agricultural employment (possibly the same agent who completes the ETA 223 form).

* The state office of Migrant Education (often attached to the department of education or superintendent of schools).

Although more expensive and time-consuming, it is also suggested that Local Review Forms be sent to the following offices in every county in the state:

* County extension office.

* Local state employment/Job Services office.

.* Local school district.

The Review Form should be sent by mail and, if possible, accompanied by a stamped, self-addressed envelope, to encourage return of completed forms. Any reviewers requesting a detailed explanation of the county classification method-ology should be sent the paper prepared on this subject. Approximately one week after the deadline for the Forms, a follow-up reminder should be mailed to those who have not returned forms or these individuals should be telephoned.

D. Analysis of Local Review Forms

Returned Local Review Forms should be grouped by state. The first examination of these Forms should look for counties which two or more reviewers indicate are misclassified. For such counties the rationale and data presented by the reviewers in support of misclassification should be carefully considered in light of what is known about the state and county make-up and the data sources used to classify the county. This examination should look at analysis errors or data quirks which might have resulted in misclassification or reasons why the data offered as evidence by the local reviewers might be in error. Whenever there is a question on comments made by the local reviewer he/she should be called. The forms should then be examined for counties identified as misclassified by only one local reviewer and the same consideration process employed.

The results of this examination must be thoroughly documented, particularly where local review comments result in a reclassification of a county. The report presenting the final classification of each county should include this narrative. Documentation as to why local review comments did not change a county classification should be kept on file in case questions are raised concerning this issue.

The best method to analyze review forms would be to employ two or more researchers in examination of local reviewer comments. Each researcher could independently write his opinion of the information presented. These opinions could then be compared and provide documentation on reclassification or non-reclassification of counties. It might be most useful if the independent researchers had various involvement in development of the preliminary county classifications; e.g., the contractor who gathered the source data, the statistician who applied the discriminate analysis process to the data and the consultant who coordinated the entire methodological process.

Example of

Local Review Form

and Instructions

INSTRUCTIONS FOR COMPLETION
OF THE LOCAL REVIEW FORM

Please help the Office of Migrant Health by checking the following designation of each county in your state as a high or low migrant and seasonal farmworker (MSFW) impact area. The definition used by the Office of Mgrant Health for a MSFW is as follows:

The term "migratory agricultural worker" means an individual whose principal employment is in agriculture on a seasonal basis, who has been so employed within the last twenty-four months, and who establishes for the purposes of such employment a temporary abode.
The term "seasonal agricultural worker" means an individual whose principal employment is in agriculture on a seasonal basis and who is not a migratory agricultural worker.
[This definition includes] migratory agricultural workers, seasonal agricultural workers and the members of the families of such migratory and seasonal workers.

A "high MSFW impact area" is defined as one which contains 4,000 or more MSFWs, including family members, at any one time during the year. A "low MSFW impact area" has less than 4,000 MSFWs at peak presence.

Each county listed on the Local Review Form is checked as a high or a low impact area. Next to the designation, there is room for you to agree, disagree or indicate your uncertainty concerning this designation. For each county you feel is misclassified, please indicate why. Your reasons should be substantiated with quantitative data, as much as possible, which will enable us to review your comments in light of the quantitative data we have used to classify the county.

In performance of this review, we ask that you keep in mind the county geographic barriers. Many counties group to form agricultural areas which are definite high MSFW impact areas, but each county considered separately may not be a high impact area. In actual use of county impact classifications, the Office of Migrant Health will group counties into health care service areas which will more closely correspond to agricultural production areas. However, for purposes of this present impact classification process, it is more productive to work on the smallest uniform geographic unit -- counties. Also, please remember when you review the designation of counties as high or low impact areas that we are concerned with both migrant farmworkers and seasonal farmowrkers as well as members of their families.

To help keep this review process on schedule, please return the compelted Local Review Form by _____ to the address lsited at the bottom of the form. We appreciate your assistance in review of these county designations and are sure the entire classification process will benefit through your help.

Example of Local Review Form

Name_____   Title_____

Agency/Organization_____

Address_____

Telephone_____

State: WASHINGTON

County Classifications

County	Designation High Impact	Low Impact	Agree	Disagree	Uncertain
Clallum		X	___	___	___
King	X		___	___	___
Whatcom	X		___	___	___
Walla Walla	X		___	___	___
. . . .			___	___	___

Reasons for Classification Disagreement
(Please provide quantitative data)

County:_____   Reasons:_____

_____

_____

_____

_____

County:_____   Reasons:_____

_____

_____

_____

_____

County:_____ Reasons:_____

_____

_____

_____

County:_____ Reasons:_____

_____

_____

_____

_____

County:_____ Reasons:_____

_____

_____

_____

_____

County:_____ Reasons:_____

_____

_____

_____

County:_____ Reasons:_____

_____

_____

_____

_____

- 128 -

County:_____ Reasons:_____

_____

_____

_____

_____

County:_____ Reasons:_____

_____

_____

_____

_____

County:_____ Reasons:_____

_____

_____

_____

_____

Please Return This Form To:

_____

Office of Migrant Health

_____

_____

_____

Telephone:_____

VI. OTHER CONSIDERATIONS IN DESIGN OF THIS

METHODOLOGY -- DISCUSSION OF FURTHER WORK NEEDED

At the conclusion of this methodological development, counties in ten states will be classified as high or low MSFW impact. (Information through Phase II, to the point of data gathering relative to discriminate analysis, will be assembled for two additional states.) Forty states will still remain which need their counties classified as high or low impact.

From the beginning of this methodological development, the OMH has wanted to devise a means for Migrant Health Clinic staff to gather a base of information about individual clinic service area populations which could be periodically updated. However, it seems that the system which has been designed to estimate the target population for the Migrant Health Program does not provide demographic data or even enumeration of the population. It is rather technical in that it requires sophisticated computer manipulation of data, and it defines high and low impact counties relative to state-wide predictors making its developmental unit the state not an individual clinic service area.

The status of high and low impact counties can be periodically verified through the use of this methodology; however, there is no guarantee that new information from the same data sources will result in reliable predictors of high and low impact counties. This means that to update the classification of counties with total assurance it would be necessary to regather data and develop new predictors on a state-wide basis.

In other words, Migrant Health Clinic staff may not be qualified to perform this method, but more important, this method may not supply clinics with the information they need to plan effective service delivery. The OMH has a desire to know more about their target population, particularly a better enumeration, documentation of migrant travel patterns, and other demographic data such as

household size and make-up including age. It would also be useful to have detailed information on the health status and health education knowledge/needs of those the clinics are designed to serve. The OMH has had to settle for the present data collection scheme, which will only define high and low MSFW impact counties, because of a legislative need for some type of target population estimate. In truth, a more exacting survey is needed to gather information crucial to planning, but the stumbling block is a lack of funds.

The OMH requires its clinics to perform a MSFW needs assessment every year as part of their grant application. The clinics perform various tasks and gather varying data to fulfill this requirement, but they are at a great disadvantage in knowing if the information they collect in any way reflects the true characteristics and conditions of their entire target population. They are not at fault in performance of this assessment because there is no methodology currently available which presents a scientific and reliable system to gather data on MSFWs. Some of the reasons for lack of this methodology are discussed in one of the papers prepared for Data Expert Conference participants, as presented in the Phase III section of this Final Report. Although the OMH and every other organization which deals with MSFWs (Migrant Education, the U.S. Department of Agriculture, the U.S. Department of Labor, the Immigration and Naturalization Service, numerous state and local government service providers and planners, and the agricultural industry -- to name just a few) need these data; specifics on MSFWs are not available. The Bureau of the Census does not collect information on this group and the Bureau of Labor Statistics specifically excludes MSFWs from its estimates.

An effective data gathering methodology has not been developed and no large scale effort is underway to rectify this lack, although such detailed information is needed by a variety of agents. The major conclusion of the participants at the Data Expert Conference is that a survey of MSFWs is the only means to

truly define the population. The consultant who has been working on this methodological development for the OMH heartily agrees. The following scheme is proposed as one means to devise and implement the necessary survey methodology.

As noted, numerous federal agencies as well as state government arms and the private agriculture industry need data on MSFWs. It seems appropriate that they should all work together on such an effort, and each agent should pick up part of the cost. The Data Expert Conference concept used by the OMH in this methodological development seemed to be a very effective means to address a difficult problem. This Conference brought together individuals from various disciplines, with different experiences and background and from all parts of the country to discuss together a methodological problem and propose a group-generated solution. The concept of experts designing a methodology should again be used to develop a survey process for MSFWs. This methodology should be on a longer developmental time frame than the current OMH effort.

The goal for the MSFW survey should be to design an appropriate method to be instituted three to five years from the present time. Each agent to benefit from the gathered data should contribute their participation and committment to the idea, as well as approximately $15,000. The funding should be placed in a central pot to be used to convene and coordinate a group of experts who will meet periodically over a six to twelve month period to develop the necessary survey methodology. A coordination agent should be responsible for selecting participants, developing a work plan, convening the groups, providing necessary assistance between meetings and assuring the methodology is completed.

The first and one of the hardest tasks of this group of experts will be to determine a survey definition of MSFWs which will meet the needs of all who want information. In addition, the developed survey method should specify exact steps and who should be responsible for local surveys, field monitor of this process and analysis of the results.

Once the survey methodology has been developed it will take the committment of all involved agencies to fund its undertaking. Those participating at the OMH's Data Expert Conference could not estimate the cost of such a survey, but it will be a considerable amount. This is one reason that initiation of data collection should be planned for three to five years in the future to allow each involved agency to institutionalize its committment to this effort.

This discussion of how reliable MSFW data can be collected may seem far removed from reality. For years, various reports and federal agencies have talked about the need for a common definition of MSFWs and the necessity to have reliable information on the make-up of this group. Until such data are available, the OMH, which first addressed the problem in its 1973 Estimate of the Target Population for Migrant Health Programs, cannot effectively plan to meet the health needs of MSFWs; and local Migrant Health Clinics cannot truly assess their service delivery. The OMH has taken the first step by convening a group of data experts to design a MSFW information gathering system and by involving other agencies in this process. The data experts are intrigued with this methodological problem, and the other agencies which need MSFW characteristics data are watching the OMH's present effort. It is time for the OMH to take the next step in acquiring the information it really needs.

The involvement of Migrant Health Clinic staff in any MSFW survey design should be carefully considered. These personnel are certainly familiar with their local service areas and could be most helpful in identifying sites to locate MSFWs for survey, as well as in encouraging the farmworker community to participate. Clinic staff, however, vary greatly in their knowledge of scientific research technique and their experience in sampling and interviewing with a standardized instrument. These personnel could be trained to play a large role in local data collection, but such training would have to be comprehensive and uniform. It would involve a great deal more than simply writing

instructions. Additionally, extensive field supervision, probably using an outside source, would be necessary to assure that the data are reliable and collected in adherence to prescribed procedures. Data analysis would most probably be handled by an agent other than the clinics.

Involvement of Migrant Health Clinic staff in this survey design to collect data would require an extensive time commitment from these individuals and would certainly affect service delivery by drawing away health care-related personnel. Some arrangement would need to be made to reimburse the clinic for staff time lost from actual clinic services or some compensation made in submission of program indicators to PHS. Most clinics are jointly funded through Community as well as Migrant Health which would require arrangements to be made with other divisions of PHS if clinic staff are to be used in a local data gathering effort. In addition, if the design depended totally on Migrant Health Clinic staff to conduct surveys in their service areas, a system would need to be found to obtain MSFW information on counties not included in existing service areas.