

Methodologies for Estimating the Migrant Population
Susan Gabbard; Aguirre International
August 5, 2005

I. Introduction

Aguirre International submits this paper to the Legal Services Corporation (LSC) to provide LSC with analytical information regarding methods for estimating the population eligible for migrant legal services. The paper discusses the principle methodologies for estimating the migrant population used by federal agencies that provide targeted services to migrant and seasonal farmworkers and their families. Other major federal programs include Migrant Health, Migrant Education, the National Farmworker Job Training Program, and Migrant and Seasonal Head Start. After describing the primary methodologies used to estimate the migrant population for these programs, the paper assesses the relative benefits of the Demand for Labor approach, on which the current distribution of migrant legal services funding is based, and an alternate approach, the Labor Expenditure methodology.

II. Current Methods Used For Population Estimates

Every federal migrant program needs information on the population that it serves for program planning and implementation purposes, as well as for understanding performance information. However, programs that serve migrant and seasonal farmworkers (MSFW) have particular difficulties obtaining this information. Due to the way that the federal government collects farmworker statistics and the multiple definitions of migrant and seasonal farmworkers used in federal programs, each program must calculate its own estimate of its eligible population. For state level distributions, the task is more complicated as migrant farmworkers move about the country and must correctly be attributed to the multiple states that serve them.

In addition, methods of estimation must follow certain principles in order to achieve fairness. These include:

- Developing a readily understandable formula;
- Using the best methods for approximating service population definitions;
- Obtaining the best available data; and,
- Ensuring that the data that are comparable for all states.

Sometimes these last two principles conflict in that data quality and availability vary across states. So, while some states have more extensive or better data, these data are hard to use unless there is comparable data for all other states.

Martin¹ was the first to commit to paper the basic principles of estimating farmworker populations and outlining the major data sources. While some of the data sources have changed since the book was written and though the methods have become more sophisticated, the basic principals are still the same.

¹ Martin, Philip, Harvest of Confusion: Migrant Workers in U.S. Agriculture. Westview Press. Boulder, CO. 1988.

“Bottom-up” measures start with local data and then build to national estimates. “Top-down” measures generally start with national level data and then work down to local estimates. Not surprisingly, bottom-up measures tend to have stronger local estimates and run into issues of comparability and fairness at higher levels of aggregation, such as states and the nation. On the other hand, top-down estimates tend to be better at the national, regional and state level and have larger errors at local levels such as counties.

Agencies serving MSFWs have taken a variety of approaches to estimating service populations. Recent data releases, particularly the release of the 2002 Census of Agriculture, have sparked interest in updating estimates of farmworker service populations.

Migrant Education: Migrant Education uses a bottom-up approach to estimating their service population. There is an extensive network of migrant education recruiters working in local areas identifying students that meet the Migrant Education eligibility definition. State totals are aggregated from local school district totals which in turn rely on the numbers of eligible students identified by local recruiters. State reports are currently reviewed and aggregated by Aguirre International under contract to the Department of Education and published in a Consolidated State Performance Report.

As is inherent in bottom-up numbers, areas with more or better recruiters locate a higher proportion of eligible students. The aggregate numbers reflect these inconsistencies in local estimates. These inaccuracies can perpetuate or exacerbate funding gaps that shortchange underserved areas, especially when the eligible populations in these areas may be increasing.

Migrant Health: Migrant Health produces population estimates for some states but has no overall numbers for its eligible service population. Migrant Health also uses a bottom-up approach based on an adjusted Demand for Labor model developed by Larson Assistance Services. The model is based on acreage of crops in production in each county and labor demand models for each crop (usually collected from state and county extension agents); local estimates are then reviewed by individuals knowledgeable of farm labor conditions. The estimates are very labor intensive and expensive, costing \$30,000 per state. In addition, due to the labor intensive nature, the estimates have been developed sequentially, with estimates for 10 states coming out over a two year period.² Ten states were funded by Migrant Health, and the bureau has no plans to have estimates for all states. A few of the remaining states have found funds and contracted independently with Larson Assistance Services.³

The Larson estimates are a good attempt to use a comparable data driven method to produce local estimates. However, while some data is available consistently across counties, other data vary in quality and availability. In addition, the use of local knowledge provides another variation in quality. As with other bottom-up estimates, comparability of estimates across counties and higher aggregations such as states is compromised by variations in the methods and sources used. Areas with more states and where individuals are more involved in the estimate

² These estimates can be accessed at: <http://bphc.hrsa.gov/migrant/Enumeration/EnumerationStudy.htm>.

³ A representative example of these can be accessed at www.casaoforegon.org.

review have better estimate quality than areas with sparse information and/or few individuals to provide local knowledge.⁴

National Farmworker Job Training Program (NFJP): In the late 1990s the Department of Labor's Employment and Training Administration (ETA) revised its estimation formula which is also its funding formula. Previously, ETA had used Census data. However, the use of Census data came under fire after research showed serious undercounts of farmworkers in the 1990 Census⁵, resulting in a 1994 letter from Everett Ehrlich, Administrator, Economics and Statistics Administration, U.S. Department of Commerce to federal agencies stating that decennial census data were not appropriate for allocating funding for programs serving migrant and seasonal farmworkers.

ETA probably has the most rigorous top-down formula. The ETA top-down estimate avoids problems with local estimates by stopping at state level estimates and not doing any county or local estimates. Currently, it calculates estimates of each state's share of the MSFW population based on a formula that includes data from the Census of Agriculture, USDA's Quarterly Agricultural Labor Survey and the National Agricultural Worker Survey (NAWS).

The formula was designed by a committee at DOL based on input from experts and reviewed by Professor Philip Martin of UC Davis. It was published for comment in the Federal Register and ETA staff and Aguirre International staff gave a series of workshops on the formula to ETA grantees before the formula was implemented. Aguirre International has calculated ETA MSFW allocation three times since the formula was developed, most recently in 2005.

Migrant and Seasonal Head Start: In the mid-1990s Professor Philip Martin collaborated with Aguirre International to devise a formula to estimate MHS slots that was a top-down estimate very similar to the ETA estimate. It relied on the 1992 Census of Agriculture, the Quarterly Agricultural Labor Survey and the National Agricultural Survey.

When the 1997 Census of Agriculture came out, there was a desire to update the estimate. In addition, the program now included seasonal farmworkers in its definition and there was a desire for an estimate of children, not the awkward use of slots. In 2001, Xtria and Child Trends estimated the number of MSHS children per state broken down by migrant and seasonal farmworkers.⁶ The Xtria estimates relied heavily on the Aguirre formula. However, Xtria made two changes that rendered the estimates problematic for researchers. First, Xtria understated the size of the farm labor population by not including families employed by farm labor contractors. Secondly, there was insufficient information on how Xtria moved from slots or FTEs to a child

⁴ For a more thorough critique of limitations of the demand for labor model see Phillip Martin's analysis in a 2001 issue of *Rural Migration News* (http://migration.ucdavis.edu/rmn/more.php?id=490_0_3_0).

⁵ (e.g., Kissam, Edward, Susan Gabbard and Philip Martin . "The Impact of Migrant Travel Patterns on the Undercount of Hispanic Farm Workers", in *Bureau of the Census 1993 Research Conference on Undercounted Ethnic Populations Proceedings*, U.S. Department of Commerce, October 1993

⁶ Clatyon, Douglas and Richard Wertheimer. Descriptive Study of Seasonal Farmworker Families, Report submitted to Administration for Children and Families, Head Start Bureau by Xtria/Child Trends 2001.

estimates and it appeared that there may have been double counting. It is possible that these two factors would cancel out but there is no information on the size of each effect.

III. Comparing methods of estimation

Aside from Migrant Education's use of administrative data, the two most popular methods of estimating MSFW populations are the bottom-up Demand for Labor Method and the top-down Labor Expenditures Method. The two methods have some basic similarities but also some important differences. Both methods begin by adjusting data to obtain hours worked in agriculture for a particular state. The Demand for Labor Model adjusts USDA acreage data using labor per acre factors culled from research and industry studies and experts as well as information from local individuals. The Labor Expenditures Model uses USDA labor expenditure data from the Census of Agriculture and divides it by regional wage data collected by USDA in its Quarterly Agricultural Labor Survey. Both methods then rely on NAWS data to move from hours worked per state to the final numbers or shares of workers per state. The Demand for Labor model also uses administrative data collected on farmworker job training clients from 1994-1998. The Labor Expenditures Model uses only adjustments that are formula based while the Demand For Labor model then goes on to allow input from local knowledgeable individuals to modify its formula outputs.

The Demand for Labor Model focuses on identifying the number of migrants in a particular county or state. In many cases it probably provides better county level estimates than the Labor Expenditure Model, but it is very labor intensive and has serious problems with data comparability at both the county and at higher levels of aggregation. The method multiplies acreage by hours per acre and divides by maximum hours worked per day and then multiplies this amount by the season length. ($\{[Acreage * \text{hours per acre}] / [\text{maximum hours worked per day} * \text{maximum hours worked per day}]\} * \text{season length}$.) Obtaining information on season length and hours per acre is very labor intensive. In addition, much labor is expended determining the types of crops and cropping methods used in each county and obtaining labor per acre estimates for those crops and methods. Moreover, methods and data sources can vary significantly from county to county. These inconsistencies are even greater from state to state.

Consider some simple but representative examples. Processing tomatoes and fresh tomatoes have different labor requirements, and how fresh tomatoes are staked further affects labor needs. Grapes can include wine grapes, table grapes or raisin grapes, each have different labor requirements. Further, a key question with respect to raisins, is whether they were dried on the vine or dried on paper trays on the ground? Similarly, was lettuce field packed or harvested in bulk? Each production method requires a separate coefficient for labor demand. In addition, expert advice must be sought to understand local cropping patterns – e.g., what percent of raisins are currently dried on the vine versus dried on paper trays. The quality and recency of the information on each method and the quality of the expert advice affect the outcomes. Where advice is good, the demand for labor method may achieve good estimates of hours worked per county. However, when the quality of information is uneven across counties in a state, or across states, then the quality of the estimates varies. In areas of intensive agriculture and for the most important crops, the information is likely to be available and of good quality. For smaller and

specialty crops and for areas where agriculture does not dominate the economy, the information is likely to be less available.

The Demand for Labor estimates use information from NAWS and job training clients to adjust for multiple jobs per worker and to obtain information on the characteristics of workers in the states. But no adjustments are made for the times migrants may reside in their home base state or elsewhere and are not working in agriculture, nor are estimates made to allocate migrant's time across multiple states. These adjustments have not been made not because they cannot be; rather, they are not in the scope of the Demand for Labor model as currently implemented. The purpose of the analyses conducted in 10 states for Migrant Health was not to estimate unduplicated shares but rather to provide county level estimates for Migrant Health grantees.

A more important and overriding concern is the unevenness of the methods and data used in the demand for labor model. Because of the intensive and local nature of the estimations, a variety of local numbers are needed. The review process allows experts and grantees to review the numbers and point out additional local criteria or numbers that might affect the calculations. Again, this may improve the quality of some local estimates but it also raises issues regarding the reliability and validity of estimates across states. This makes the process uneven and potentially unfair. Intensive agricultural states with many resources may get better reviewers than states with fewer resources.

Finally, because it is so labor-intensive, the Demand for Labor model is very expensive. As indicated above, the studies conducted for the Migrant Health Program cost some \$30,000 per state.

The Labor Expenditures model is fundamentally a top-down model focusing on what is the proportional share of each state in the national farm labor population for a specific migrant service program. The model uses uniform, national data sets that are consistently applied across states. These data sets are based on scientific samples. USDA data are used for both labor expenditures and wage data. Both data sources belong to the National Agricultural Statistical Services, which provides information on the accuracy of its estimates. One concern with the labor expenditures data is that they vary by state due to the required state benefits. States that require unemployment insurance and workers compensation insurance will have higher labor expenditures than states with the same labor hours that do not require these programs for agricultural employees. In addition, there is concern that the labor expenditures include some ineligible employees in the labor expenditures and that this effect varies by state. Little information is available about the magnitude of these effects or their impact across states. Methods can be developed, however, that would make consistent adjustments among states for these factors.

The Labor Expenditures model uses only NAWS data to move from hours to state shares. The adjustments in the Labor Expenditures model include adjustments for migrants who divide their time among states as well as for states where farmworkers spend time when they are out of work. One problem common to both of these populations is that the NAWS does not exactly replicate service population definitions used by agencies and so the match is inexact. Adjustments can be made, such as those made to the NFJP formula, that can ensure results are accurate and fair.

Further, this is less problematic when the purpose is to ascertain the distribution of the migrant population than it is when the goal is to estimate the size of the migrant population (as is the case with the Demand for Labor formula).

Thus, the Labor Expenditure estimates are based on methods and data sets that are far more uniform, consistent, and reliable than those used by the Demand for Labor approach. Additionally, preliminary Labor Expenditures estimates are reviewed for consistency and quality. Apparent anomalies are examined in reference to the available data. If adjustments are made, they are made by the agency based on uniform revisions to formulas. Input from stakeholders has proved useful in the development of these revisions, but these modifications must be applied in a uniform and consistent manner, not on a case-by-case basis for individual states.

Finally, the Labor Expenditures estimates are less labor intensive and therefore cheaper and faster than Demand for Labor estimates. They can usually be completed in less than six months and can cost from \$50,000 to \$100,000 for all states depending on the level of adjustments and reporting required.

The following chart highlights the major strength and weaknesses of the Demand for Labor and Labor Expenditure models.

Exhibit 1: Comparing Demand for Labor and Labor Expenditures Methods

	Demand for Labor	Labor Expenditures
1. Hours worked in farmwork	Derived from USDA acreage data multiplied by labor per acre factors obtained from studies for each crop and method of production	USDA labor expenditure for hired and contract labor divided by regional wages from USDA Quarterly Agricultural Labor Survey
Issues:	Quality of labor per acre coefficients varies. Not all methods may have reliable or current labor per acre factors. Labor per acre may vary in ways not captured by studies as growers adopt new methods. Requires making decisions about cropping patterns -- are acres of tomatoes processing, on poles. Increases effort and cost of estimate.	Labor expenditures vary by state with differences in required benefits such as workers compensation, unemployment insurance. Should be easy to adjust for. May includes some ineligible employees and this may vary across states. May possibly be able to be adjusted.
2. Hours worked per worker	Relies on NAWS data factors and other factors to turn hours into workers with specific eligibility characteristics.	Relies on NAWS data to determine proportion of hours that are eligible for services.
Issues:	NAWS and other data provide only an approximation of the eligible population.	NAWS provides only an approximation of the eligible population.
3. Allocate migrants across states.	No adjustments made but could adjust workers using NAWS.	Workers allocated to home base states where they do not work and migrants time pro-rated across multiple states.
Issues:	No cross state adjustments are made for workers who spend time in more than one state. Workers who spend out of work time in a homebase state are not counted in that state unless they work there as well.	
4. Additional adjustments	Estimates reviewed by local knowledgeable individuals in each state who recommend adjustments to the count. Adjustments could be eliminated but might be integral to quality.	Estimates reviewed for quality and consistency. Formulas and factors adjusted only if directed by agency.
Issues:	Quality of review varies by state depending on reviewers. Adjusting in this method means method is not consistent across states, possible unfairness if quality of review and reviewers not the same in each state.	
5. Timeliness	Because it is labor so labor intensive, estimates for all 50 states could potentially take two year or more to produce.	Approximately six months.
6. Cost	At \$30,000 per state, costs would be \$1,500,000 for all 50 states.	Actual cost depends on scope but once a formula is developed, ongoing estimates could be produced for less then \$100,000
Issues:	Method is very labor intensive and thus expensive.	Use of national data sets reduces costs.