BY THE U.S. GENERAL ACCOUNTING OFFICE

Report To The Chairman Of The Subcommittee On Immigration And Refugee Policy Of The Committee On The Judiciary United States Senate

Problems And Options In Estimating The Size Of The Illegal Alien Population

Illegal aliens are of concern to the Congress not only because of their illegal status but also because they may aggravate employment and community resource problems. As the Congress considers its response to the presence of illegal aliens in this country, accurate estimates of the size and growth of this population would be useful for deciding on policy options and for evaluating policy effectiveness. However, presently available estimates are imprecise and insufficiently reliable.

GAO reviewed previous estimates of the size of the illegal alien population as well as methods that have been used to estimate other populations that are difficult to count. GAO found that acquiring precise estimates of the illegal alien population with these methods requires better information on immigration than is presently available.

GAO presents for congressional consideration three alternative ways of acquiring information relevant to policymaking on illegal aliens. In assessing the merit of these alternatives, the Congress should weigh the extent of its concern for reliable narrow-ranged estimates against the significant expenditure of resources that would be required.

GAO/IPE-82-9
SEPTEMBER 24, 1982

Request for copies of GAO reports should be sent to:

U.S. General Accounting Office
Document Handling and Information
Services Facility
P.O. Box 6015
Gaithersburg, Md. 20760

Telephone (202) 275-6241

The first five copies of individual reports are free of charge. Additional copies of bound audit reports are $3.25 each. Additional copies of unbound report (i.e., letter reports) and most other publications are $1.00 each. There will be a 25% discount on all orders for 100 or more copies mailed to a single address. Sales orders must be prepaid on a cash, check, or money order basis. Check should be made out to the "Superintendent of Documents".

UNITED STATES GENERAL ACCOUNTING OFFICE
WASHINGTON, D.C. 20548

'ITUTE FOR PROGRAM
EVALUATION

B-209064

The Honorable Alan K. Simpson
Chairman, Subcommittee on Immigration
and Refugee Policy
Committee on the Judiciary
United States Senate

Dear Mr. Chairman:

In your October 26, 1981, letter you asked that we review the methodologies used in measuring hidden populations for their potential in improving estimates of illegal aliens. As requested, we briefed your staff on our major findings on April 26, 1982. The report, "Problems and Options in Estimating the Size of the Illegal Alien Population," reviews the previous estimation attempts, describes the problems attending the use of alternative estimation methods, and presents options for acquiring improved policymaking information in the illegal alien area.

We sought comments on the report from the U.S. Department of Justice. Their comments and our response are included in the report as appendix V.

Copies of this report are being sent to the Chairman of the Subcommittee on Immigration, Refugees, and International Law of the Committee on the Judiciary, U.S. House of Representatives; to the Department of Justice; and the Immigration and Naturalization Service. Copies will also be made available to other interested parties who request them.

Sincerely yours,

Eleanor Chelimsky
Director

GENERAL ACCOUNTING OFFICE
REPORT TO THE CHAIRMAN
SUBCOMMITTEE ON IMMIGRATION
AND REFUGEE POLICY
COMMITTEE ON THE JUDICIARY
UNITED STATES SENATE

PROBLEMS AND OPTIONS IN ESTIMATING THE SIZE OF THE ILLEGAL ALIEN POPULATION

D I G E S T

Illegal aliens are foreign nationals who have entered the country (1) without immigration documentation or inspection or (2) with fraudulently obtained immigration documents or (3) legally but later violated the terms of their entry documents. Knowing their numbers is important not only because of their illegal status but also because they are considered to be aggravating employment and community resource problems as well as themselves suffering from social and health problems caused by their "sub rosa" status. However, available estimates of the size and growth of the illegal alien population are unsatisfactory for policymaking. The estimates, derived in several different ways, vary greatly and are not very defensible.

Accurate estimates of population size for other hidden populations--that is, those that are difficult to count--have been made. Prospects for applying any of those methods to the estimation of the size of the illegal alien population are poor, however, because our knowledge is limited. Nonetheless, some methods do exist for providing information about illegal aliens that is relevant to the formulation of policy. In this report, GAO presents for congressional consideration three ways of acquiring such information.

GAO undertook this review at the request of the Chairman of the Subcommittee on Immigration and Refugee Policy and the Senate Judiciary Committee. Accordingly, in analyzing methods used previously to estimate the size of the illegal alien population as well as other hidden populations, GAO's objective was to determine whether any existing methods might produce a valid and reliable estimate of either the national resident illegal alien population or the annual flow of illegal aliens into the United States. GAO reviewed the research on legal and illegal immigration,

Tear Sheet.

interviewed private researchers as well as officials of the Immigration and Naturalization Service and the Bureau of the Census, analyzed methods used to estimate other hidden populations for their data requirements and assumptions, and, finally, evaluated their utility in estimating the stock and flow of the illegal alien population. (pp. 2-5, 15-18)

ESTIMATES OF THE ILLEGAL ALIEN POPULATION

Estimates of the resident illegal alien population range from 1 million to 12 million, with the most widely accepted range being 3.5-6 million. The most frequently cited estimate of the number who enter illegally each year is 500,000. GAO found no single previous estimate of either the national illegal alien population or its annual flow to be both valid and reliable. Current estimates stem from incomplete or questionable data bases or untested or demonstrably incorrect assumptions or are restricted to a subgroup of the illegal alien population. (pp. 10-11)

METHODS USED TO ESTIMATE OTHER HIDDEN POPULATIONS

GAO identified in the literature three general approaches to estimating the size of other hidden populations. Two require the identification of population members and place special constraints on sampling procedures. The third avoids the identification problem but requires a solid understanding of the targeted population. For these estimation approaches to yield defensible estimates, a great deal has to be known about the characteristics of the population being studied and about the ways in which that population arises and is distributed. GAO's review of the available information on illegal aliens, however, indicated that they constitute a highly diverse population whose behavior and composition are unknown and in flux. Therefore, the size of the illegal alien population cannot be accurately estimated at this time. (pp. 12-14)

POSSIBILITIES FOR IMPROVED POLICYMAKING INFORMATION

Producing reliable estimates from available methods would require an extensive, expensive,

and time-consuming research program. However, a nationwide estimate of the number of illegal aliens is not the only useful form of information about the population. Estimates of change in both the size of the illegal alien population and its flow into the country may be sufficient for some policymaking purposes. In this report, GAO describes one method that could provide estimates of relative change (that is, change expressed in percentages)--the "multi-indicator method." It assumes that some indicators of the illegal alien population may be subject to too much error to serve individually as reliable estimators but that as a group their errors may counterbalance each other. (pp. 19-23)

Another method GAO discusses could yield accurate estimates of various subgroups of the illegal alien population. Some attempts to measure specific subgroups have been successful--for example, aliens who overstay the terms of their temporary visas have been counted by comparing the dates on the relevant forms. The usefulness and accuracy of such information depend on the specific subgroup in question and have to be assessed carefully case by case. (pp. 23-24)

Estimating the illegal alien population as a whole requires more specific information about that population than presently exists. The understanding of illegal alien behavior could be expanded by ethnographic research. Surveys would also provide important prevalence and frequency information. To assess the overall merit of undertaking such a research program, however, the Congress should weigh the extent of its concern for a reliable, narrow-range estimate against the expenditure of the significant resources that would be required at both the national and the local levels. (p. 24)

AGENCY COMMENTS

The Department of Justice concurred with the major findings of the report. The final version of the report incorporates changes suggested by the Department, whose comments are included as appendix V.

Tear Sheet.

Contents

ABBREVIATIONS

CENIET	Centro Nacional de Informacion y Estadisticas del Trabajo
CPS	Current Population Survey
DOJ	U.S. Department of Justice
GAO	U.S. General Accounting Office
INS	U.S. Immigration and Naturalization Service
IRS	Internal Revenue Service
SSA	Social Security Administration

CHAPTER 1

INTRODUCTION AND ISSUE DEFINITION

Illegal aliens are foreign nationals who have (1) entered the United States without immigration documentation or inspection, (2) entered with fraudulently obtained documents, or (3) entered legally but later violated the terms of their entry documents, as by staying beyond the expiration dates of their documents or by taking unauthorized employment. Illegal aliens have also been termed "illegal migrants," "undocumented workers," and "deportable aliens." The last term is preferred by the U.S. Immigration and Naturalization Service (INS). In addition to the fact that illegal aliens constitute a large number of people who are in violation of the law, their presence in the United States raises several related policy issues.

The U.S. Select Commission on Immigration and Refugee Policy has expressed its concern that illegal aliens may displace U.S. citizens and immigrants who are legal permanent residents from jobs and depress local wages. The Commission found no evidence that illegal aliens drain government and community services without contributing to the tax base. The Select Commission was also troubled about the development of an "underclass" with unequal rights and unequal participation in U.S. society. 1/ As members of a "sub rosa" population, illegal aliens are expected to be highly susceptible to poor working conditions, criminal victimization, and poor health, since they avoid making contact with public officials who provide services in these areas.

Estimates of the number of illegal aliens are a necessary statistic for policymakers. Accurate estimates are required for making sound, relevant immigration policy; judging the effects of deterrence and enforcement policies; estimating the pull of illegal aliens on welfare, education, and other items in State and local budgets; assessing the need for special labor market strategies; and making realistic allocation of resources to INS. In addition, accurate estimates are useful to INS in assessing the effectiveness of its alternative management and apprehension strategies.

At present, however, neither the number of illegal aliens currently residing in the United States nor the number entering the country annually is known. In April 1981, the Select Commission on Immigration and Refugee Policy published the estimate

---

1/ In 1979, the Select Commission on Immigration and Refugee Policy was created by Public Law 95-412 to study and evaluate existing laws, policies, and procedures governing the admission of immigrants and refugees to the United States. Its final report was published April 30, 1981.

that between 3 million and 6 million foreign nationals reside illegally in this country. This estimate was based primarily on a review by Bureau of the Census staff members in January 1980 in which they "cautiously speculated" the existence of 3.5 million to 5.0 million illegal residents in 1978 (see Siegel, Passel, and Robinson, 1981). 1/ Several other attempts have been made to estimate this population, but the range of the estimates and their statistical reliability are not considered satisfactory, as we noted in our report Number of Undocumented Aliens Residing in the United States Unknown. (GAO, 1981)

The Congress is currently considering various strategies for resolving some of these issues. The Chairman of the Subcommittee on Immigration and Refugee Policy of the Senate Judiciary Committee requested us to assist in the deliberations regarding these strategies by reviewing and analyzing the methods that are available for estimating the size of the illegal alien population. (The request letter is printed in appendix I.)

OBJECTIVE, SCOPE, AND METHODOLOGY

Our objective was to determine whether any methods have been devised for producing a valid and reliable estimate of the size of the illegal alien population in the United States. We sought and reviewed all published attempts to estimate the population at the national level in the last decade. These estimates include both the number of illegal alien residents (that is, the stock of the population) and the number of illegal aliens entering the country annually (that is, the yearly flow of the population) and are listed in appendix II. In addition to identifying these studies and other published estimates, we interviewed a number of officials and other experts in immigration and population research, including staff members of INS and the Bureau of the Census.

We reviewed all empirically based national estimates of the illegal alien population for the reliability and validity of their data collection and analysis methods as well as for their assumptions. In order to evaluate the reasonableness of their assumptions, we reviewed survey and ethnographic research on both legal and illegal immigration.

We also looked at the procedures that have been used to estimate other difficult-to-enumerate populations to see if any of them might be applicable in studying illegal aliens. We selected an illustrative, nonexhaustive group of hidden populations--namely, criminals, heroin addicts, alcoholics, disease victims, and wildlife. We based this selection on our experience

---

1/ Interlinear bibliographic references are cited in full in appendix IV.

and judgment, taking into consideration the recent development of advanced methods of estimation in some areas. We identified these through a targeted literature search and interviews with private researchers and officials at the National Institute of Drug Abuse of the U.S. National Institutes of Health and at the U.S. Fish and Wildlife Service. We reviewed each method first for its data requirements and assumptions and then for its applicability in estimating the illegal alien population.

REPORT OVERVIEW

In the balance of this report, we present our analysis of previous estimation attempts and discuss the utility of other approaches to assessing the size of a given population in general and the illegal alien population in particular. In chapter 2, we lay out a specific critique of the methods that have been used to estimate the illegal alien resident population and the number entering the country annually. We do not present an exhaustive review of the limitations of each study, but instead we highlight the problems that are encountered in using the various estimation approaches. In chapter 3, we define the concept of "hidden population" and identify the estimation methods that have been used with populations other than illegal aliens. We discuss the procedures, data requirements, and assumptions of three such approaches. In chapter 4, we illustrate the problems of applying these approaches to estimating the size of the illegal alien population. Finally, in chapter 5 we discuss three broad courses of action that might be taken to improve the information policymakers have about illegal aliens.

Five appendixes follow the text of the report. In appendix I, we reprint the letter from the Subcommittee on Immigration and Refugee Policy of the Senate Judiciary Committee requesting this report. In appendix II, we provide a complete list of original estimates of the size of both the flow and the stock of illegal aliens made in the last decade, including the studies we review in chapter 2. We present only the original source of an estimate and a shortened form of its publication reference. Appendix III is a glossary of immigration and methodological terms used in this report. Full references for the sources in appendix II and for bibliographic citations in the report are given in appendix IV, along with other relevant works. In appendix V, we reprint agency comments on a draft of this report and our response.

CHAPTER 2

THE PREVIOUS ESTIMATES OF THE NUMBER OF ILLEGAL ALIENS

Although many different methods have been used to estimate the number of illegal aliens, all have drawbacks, stemming primarily from the fact that illegal aliens deliberately try to evade identification. This forces investigators to rely on biased sampling procedures, inappropriate proxy measures, and a variety of untested assumptions that cast doubt on the accuracy of the resulting estimates.

The illegal alien population is constantly in flux. Individuals enter and exit this population daily by making covert border crossings, by taking unauthorized employment, by failing to leave when their visas expire, by dying, and as a result of INS decisions. Policymakers need to know the size of the population at particular moments and the volume of migration over periods of time. Although illegal immigration has been of concer since the introduction of immigration curbs in 1875, deliberate attempts to estimate the population size reliably were not made until the 1970's.

THE DATA BASE FOR THIS REVIEW

Over the past decade, 13 studies have provided 20 estimates of either the stock or the flow of illegal aliens. In addition, 8 other estimates have also been made public. We have listed these 28 estimates, with their year and source, in appendix II. The sources are varied and include the Immigration and Naturaliza tion Service, the Bureau of the Census, private researchers, and the Mexican government. The estimates of the national stock of illegal aliens range from 1 million to 12 million; estimates of the annual flow range from 0.8 million to 1.5 million. Some are completely speculative (including the 12 million figure), while others reflect the use of statistical analyses or survey methods and represent large expenditures of time and resources.

In 1980, Siegel, Passel, and Robinson, staff members of the Bureau of the Census, made a critique of illegal alien populatio estimates for the Select Commission on Immigration and Refugee Policy under the title "Preliminary Review of Existing Studies of the Number of Illegal Residents in the United States." Neith using new data nor reanalyzing the old, they judged the quality of pre-1980 estimates and "cautiously speculated" that 3.5 milli to 5.0 million illegal aliens were residing in the United States in 1978. In its final report in 1981, the Select Commission sug gested the figures 3.5 million to 6.0 million illegal residents, a range that has been cited frequently over the past year. The estimate of the annual flow of illegal aliens that is most fre- quently given is 0.5 million, "although it has no empirical basi (Keely, 1982, p. 1).

In this chapter, we review 12 of the 13 studies (omitting one that replicated a method on a new data set) and one uncompleted study (under contract with INS in 1977). We highlight the major problems in their estimation approaches, without attempting to present all the limitations of each study. We include no studies providing subnational estimates.

ESTIMATES OF THE NUMBER OF ILLEGAL ALIEN RESIDENTS

A major problem in identifying illegal aliens is that the individuals who have that status are reluctant to admit to it. This has frustrated all attempts to estimate the stock of illegal aliens. The weak point of the following three studies is the method by which they tried to identify the population.

Study 1

The Mexican government conducted the National Survey of Emigration in December 1978 through January 1979, in which the sampled households were asked whether any of their members over age 15 were currently in the United States and working or looking for work regardless of their legal status. (CENIET, 1979b) The study concluded that an estimated 400,000 Mexicans were temporarily residing in the United States and therefore likely to be illegal aliens. However, this is likely to be an underestimate of even this subpopulation of illegal aliens because of probable underreporting of illegal migrants, children, and entire families that had moved.

Study 2

The Mexican government surveyed a sample of illegal aliens whom INS had returned to the Mexican border. (CENIET, 1979a) They were asked about their length of stay in the United States and the number of previously forced and voluntary returns to Mexico. These responses were used with INS apprehension data to estimate that 500,000 to 1,200,000 Mexican illegal aliens resided in the country in 1977. As the Census staff noted, however, the validity of these estimates depends on how well (and how honestly) the sample of apprehended persons surveyed represents the volume and character of the movement between Mexico and the United States. (Siegel, Passel, and Robinson, 1981)

Study 3

In 1977, INS contracted with Reyes Associates for a direct survey of immigration status intended to produce a national population estimate. Reyes Associates hoped that by using local interviewers from the same country of origin, with strong guarantees of confidentiality, survey respondents would acknowledge their illegal status and the underreporting problem would thus be solved. However, cost overruns and other difficulties prevented completion of this survey,

the production of an estimate, and even INS access to the preliminary results.

One particular form of abuse of immigration law is accessible to direct enumeration. People who overstay the term of their temporary visas--commonly given to students, visitors, diplomats, and foreign crewmembers--can be counted by matching the individual departure and arrival forms submitted by all aliens who are inspected at U.S. borders.

Study 4
After attempting to match departure and arrival inspection documents, INS reported in mid-1976 that 740,000 nonimmigrant aliens had failed to depart since 1974 and that more than a million unresolved departures were left outstanding from before 1974. (Chapman, 1976) These numbers represent upper bounds for the number of overstays, since INS investigative efforts on samples of these records found that some such persons had, in fact, departed the country or were still present but in legal status. When INS has processed the backlog of these forms and has improved its record collection, this could be a reliable method of directly counting this particular segment of the illegal alien population.

The other attempts to estimate the stock of illegal aliens have followed a discrepancy model. Generally, this approach involves comparing estimates of total population size that have been based on independent sources of data. Discrepancies that are observed among them are presumed to result from the presence of illegal aliens in one estimate and their absence in another. The major problem with this approach is that illegal aliens are not likely to be completely included or excluded in any single data set.

Study 5
Goldberg (1974) compared the Mexican census count of 1970 with the number expected to result from counting births, deaths, and legal immigrants since the 1960 census. Goldberg considered the 1,597,000 difference as representing illegal migration from Mexico to the United States between 1960 and 1970. The accuracy of this method depends on two assumptions: that the underenumeration can be accurately estimated in both the 1960 and the 1970 census and that it is possible to make accurate adjustments of the 1960 count.

Study 6
Lancaster and Scheuren (1978) reported on a study that categorized households in the 1973 Current Population Survey (CPS) sample by whether or not they were represented as contributors or recipients in any of three Social Security (SSA) and Internal Revenue Service (IRS)

data files (this was referred to as the 1973 CPS-IRS-SSA Exact Match Study). A statistical procedure (log linear models employing the marginal totals from the 2 x 2 x 2 table) was used to estimate the number of people not in the SSA or IRS data systems:

"The aggregate of all eight cells (presumed to include illegal residents) is then compared with the Census Bureau's population estimate (which presumably does not include illegal residents) and the difference is taken as an estimate of illegal residents." (Siegel, Passel, and Robinson, 1981, p. 20)

The accuracy of this method depends on the accuracy of the matching of records, the completeness of the CPS coverage, the accuracy of the SSA and IRS counts, the validity of inferring total population estimates from the CPS sample, and the extent of undercoverage of illegal residents in the administrative data. The authors themselves characterized this work as exploratory.

Study 7

Warren (1981) compared estimates of the number of foreign-born respondents to the November 1979 Current Population Survey who had reported immigrating since 1970 with each of two independent INS population counts: (1) the number of legal immigrants and refugees admitted since 1970 (adjusted for expected emigration and mortality) and (2) the number of aliens reporting their addresses in January 1980 (adjusted for the expected incompleteness of registration). The roughly one million difference in each comparison gives an estimate of the number of illegal aliens counted in the November 1979 CPS. While it is recognized that some illegal aliens may not respond accurately to the CPS, this method can produce a good minimum estimate of the population size.

Study 8

Robinson (1980) developed a wide range of estimates (600,000 to 4,700,000) from analyses of trends in State death rates from 1950 to 1975. The expected numbers of deaths in selected States were computed and these figures were then compared with the actual number of deaths in those States. In each State, the observed number was higher than the expected number, and the difference was assumed to result from the enumeration of illegal aliens in death records but not in the census. Death rate data were then used to project the number of illegal aliens from the number of unexplained deaths. The major problems with this approach are that death rates may actually vary for the target States and that death rates for illegal aliens are not known and must be guessed.

ESTIMATES OF ANNUAL ILLEGAL IMMIGRATION

The number of illegal aliens entering the country each year is expressed in estimates of population flow. A distinction should be made between the gross flow of population (the total number of entries) and the net flow (the number of entries minus the number of exits). Only one study of those we reviewed attempted to estimate the net flow of population.

Study 9

Heer (1979) analyzed the change between 1970 and 1975 in the Mexican-origin population of the United States, as estimated by the Current Population Survey. The difference between the two population estimates was reduced by the estimated natural increase and by the amount of legal immigration from Mexico during this period. Using various assumptions about the CPS coverage of the population and rates of return migration, Heer estimated a range of net illegal immigration from Mexico over that five-year period of 400,000 to 1,200,000 and an average annual net immigration from Mexico of 82,000 to 232,000. The Bureau of the Census staff pointed out two major problems with this method. The sampling error of the difference between the two CPS estimates is relatively large, partly because of large sampling and response errors for this population. For example, in using this method for the period 1973 to 1977, "the results could have shown a substantial net outflow of the Mexican-origin population." (Siegel, Passel, and Robinson, 1981, p. 23) Further, they claimed that the estimates were determined primarily by the study's assumptions concerning the level of CPS undercoverage but these assumptions have no empirical basis.

Given the focus on entry, it is natural that most estimates of gross flow use the most comprehensive data on entry--namely, INS apprehension statistics. However, these data indicate only unsuccessful attempts at illegal entry. Estimating the flow requires determining how many attempts are successful--that is, how many go undetected. Therefore, several studies have focused on estimating a ratio of successful to unsuccessful entries.

Study 10

Lesko Associates (1975) employed INS apprehension statistics for Mexican illegal aliens to produce a "minimum got-away-at-entry" ratio, defined as the number apprehended away from the border divided by the number apprehended while attempting entry. Computed annually, this ratio was then increased by a "constant multiplier," giving an estimate of the ratio of successful entries to border apprehensions. The "constant multiplier" was derived from the assumed increase in the illegal Mexican population from zero in 1960 to 1,597,000 in 1970 (Goldberg's 1974 estimate). The

estimate of flow derived from this method rests upon several debatable assumptions underlying the "minimum got-away-at-entry" ratio and the "constant multiplier." For example, the "minimum got-away-at-entry" ratio mixes entries during current and previous years and also reflects differences of labor and strategy between the Border Patrol and the Investigations Division of INS. 1/

Study 11
INS attempted to estimate the number of fraudulent entries at border points, including airports. (DOJ, 1976) At airports and selected Mexican border points, special teams of immigration inspectors conducted intensive inspections of a random sample of border crossers, looking for fraudulent documents and the intention of violating the terms of valid visas. The difference between the rate of fraudulent entries thus detected and the rate detected by usual procedures was assumed to be the usual rate of successful fraudulent entry. The stability of the estimated rate is highly vulnerable to changes in personnel, procedures, and level of traffic and would require frequent updating.

Study 12
In a less direct attempt to estimate fraudulent entries, Vining (1979) analyzed data on air passenger arrivals and departures at U.S. international airports between 1974 and 1977. He estimated that the number of aliens arriving by air exceeded departures by 500,000 to 700,000. Subtracting the annual average net legal immigration by air from that net excess of arrivals yielded an annual average of 180,000 to 380,000 persons illegally immigrating by air during that period. This method cannot account for aliens departing by other modes of transportation and is completely dependent on the airlines for collecting data on departures.

Study 13
Morris and Mayio (1980) estimated the size of the illegal alien immigration flow from Mexico in 1978 by combining

---

1/Lesko Associates also estimated the stock of Mexican illegal aliens by using their estimates of annual flow. After assuming there were 1.597 million Mexican illegal aliens in this country in 1970, they then added to this figure the estimated net flow over the ensuing five-year period, assuming a 97 percent "survival" rate (higher than that found for legal aliens). In addition, a "Delphi" method was employed to estimate the entire illegal alien population. Here, a panel of experts attempted to reach consensus on an estimate of the ratio of illegal aliens to legal aliens in the United States. These ratios were then applied to the known number of legal alien residents in 1975.

adjustments of INS apprehension data. The number of Mexicans entering fraudulently was derived from the INS estimate of successful fraudulent entries at land border crossings (given in DOJ, 1976). The number of Mexican "entries without inspection" was estimated by using ratios of successful to apprehended entries from earlier subnational studies (North, 1975; Dondero, Bieber, and Forrester, 1977) to inflate the 1978 border apprehension statistics. This number was halved, because of the CENIET (1978) ratio of entries per person, to produce an estimate of the number of Mexicans entering without inspection. These two estimate were then added, yielding a total estimate of 1,075,000 to 1,735,000 Mexicans illegally immigrating in 1978. Morris and Mayio noted the unreliability of these adjustments even at the time they were originally made; their stability over the ensuing three years is even more questionable in the light of changed INS resources and strategies. 1/

SUMMARY OF PROBLEMS WITH PREVIOUS ESTIMATES

As this review has shown, a variety of problems accompany the approaches that have been used to estimate the size of both the stock and the flow of illegal aliens. The most serious are as follows.

1. Several approaches are appropriate only for a specific subgroup of the illegal alien population--aliens from a single country of origin, those who overstay their visas, residents of certain States only, and the like. This precludes generalizing the use of these procedures to the entire illegal alien populati

2. The discrepancy approach, which assumes that illegal aliens have been included (if not identified) in certain government data files but not in others, risks mistaking error in the files for the presence of illegal aliens.

3. Studies that rely on illegal aliens reporting on their own status assume that such reports are accurate. It may be, however, that they are not accurate, given the reluctance of individuals to be self-incriminating.

1/Morris and Mayio, in a procedure similar to that of Lesko Associates, estimated the stock of Mexican illegal aliens by using their estimates of annual flow. After assuming that there were 1 million Mexican illegal aliens in this country in 1970, they then added to this figure the estimated net flo over the ensuing 8 years, assuming an average length of stay of 6 months. The stock of non-Mexican visa abusers was estimated as 15 percent of the difference between the recorded number of aliens admitted and those who depart each year for the period 1951 to 1977.

4. To estimate the number of individuals crossing the border illegally from the number of border apprehensions requires adjusting for both the number of border-crossing attempts for each successful (that is, undetected) entry and the number of successful entries each person makes during a particular time period. However, since information about successful entries generally comes from small, biased samples, such adjustments must rely on untested assumptions.

The problems in measuring the stock and the flow of illegal aliens have implications for making policy decisions. For example, in the absence of a constant ratio between detected and undetected illegal entries, neither a rise nor a decline in the number of apprehensions can be interpreted as reflecting a corresponding change in the flow of successful illegal entries. The problems of identification, measurement, and estimation we have discussed in this chapter signify that formulating sound policies and evaluating their effects is especially difficult in this area. Policies seeking to reduce the number of illegal aliens successfully crossing the border, for example, would require for their evaluation a more sound knowledge of actual changes in the stock and flow of illegal aliens than we currently possess.

These problems are not limited to the topic of illegal aliens, however. The same difficulties occur in measuring other hidden populations, whether they are criminals or victims of stigmatizing disease. It is difficult to assess the success of a policy that aims to change the size of a population if one cannot be sure of the actual size of that population. In the next chapter, we review some of the methods of estimating other hidden populations.

CHAPTER 3

THREE METHODS FOR ESTIMATING HIDDEN POPULATIONS

"Hidden population" is defined here as any population whose members are difficult either to observe or, upon observation, to identify as members of that population. Examples include fish in a lake (difficult to observe), alcoholics (difficult to identify as problem drinkers), tax evaders (difficult to identify), and illegal aliens. Since the methods for estimating the illegal alien population have major weaknesses, we decided to examine methods that have been used for estimating other hidden populations to see what promise these might offer.

The literature on population estimation presents three methodological components of estimation: (1) sampling mechanisms, or procedures analysts can use to identify the population, (2) underlying population dynamics, or information on how a population arises, is distributed, and behaves, and (3) mathematical models, or the formulas that use information from the sampling and population dynamics components to generate estimates of population size. In this chapter, we present the results of our review of sampling mechanisms (which we refer to as "methods") used to estimate the size of five hidden populations in order to determine whether any might help in estimating the number of illegal aliens. In chapter 4, we demonstrate how the lack of knowledge about the dynamics of the illegal population restricts an analyst's ability to produce population estimates from data collected by these three methods. The mathematical models are discussed in the appropriate references in appendix II.

The five hidden populations we reviewed were criminals, wildlife species, heroin addicts, victims of sexually transmitted disease, and alcoholics. We based our selection of these five on our experience and judgment as well as on the recent development of advanced methods of estimation in these areas. We delineate three methods and their requirements for successful application. We refer to them as the "window," "tagging," and "indicator" methods.

Both the window method and the tagging method use the concept of reverse sampling. Sampling typically involves selecting some segment of a population (a sample) that can serve as a representative of the entire population. One of the criteria employed in determining how large a sample to select is the size of the population of interest. In some attempts at estimating hidden populations, however, this process is reversed. That is, instead of trying to estimate an appropriate sample size on the basis of a known population size, one observes some sample and bases an estimate of the total population size on the number of observations in the sample.

The window method involves counting all members of a population observed during a specific period of time and within a specific geographic area. The term "window" derives from the fact that an analyst constructs a window in time and space through which to make observations. Among the five areas we reviewed, estimating the prevalence of different wildlife species depended the most on the window method. One example of its use is the aerial observation of the coastline of Louisiana, Mississippi, and Alabama during the summers of 1974 and 1975 for the purpose of estimating the size of the bottlenosed dolphin population along that coast.

If the window method is to provide accurate population estimates, it is essential that the analyst understand how the observation period relates to nonobservation periods. Without this information, bias can be introduced into the estimate. For instance, a poorly chosen time for a window may reflect a disproportionately heavy or light number of population members.

The tagging method is based on the assumption that the frequency with which members of a population are observed is related to the size of that population. This simply means that if there are a great many criminals, we should expect to encounter criminals much more frequently than if there are very few. Probability-based statistical models allow us to use the frequency of initial and subsequent contacts with members of a hidden population to estimate the size of that population.

In the simplest case using the tagging method, if the number of ping-pong balls in a box is unknown yet every time we reach in and pull out a ball we get the same one, we can reasonably assume that there is only one ball in the box. If we pull out the same ball once in every two tries, we might conclude that there are two balls, once in every three tries that there are three, and so on. Adopting this approach requires us to be able to identify each individual so that the number of times we encounter the same individual can be recorded. The term "tagging" derives from this necessity of identifying individuals. As with the window method, the most frequent application of tagging in the five areas we reviewed was in estimating wildlife populations. It has also been used for estimating the number of criminals in the Washington, D.C., area. (Greene and Stollmack, 1980)

In order to use the tagging method successfully, we must have one of two situations. Either the chance that each member of the population will be observed must be equal to the chances of all the others or, in cases in which the chances are not equal (referred to technically in the literature as "differential capture probabilities"), the differences within the population must be known. In our example of the ping-pong ball, this means that each ball must have an equal chance with all others of being selected. If we simply throw back a ball we have pulled out and

do not shake the box, it is possible that the next time we reach in we will simply pull out whichever ball is on the top and, thereby, severely bias our estimate.

The third general approach for estimating the size of hidden populations is termed the "indicator" method. Using this approach means no attempt is made to observe directly or to count members of the population of interest. Rather, some variable that is known to be strongly related to population size is observed and used as an indicator of size. Examples include using the number of deaths caused by heroin overdose to estimate the number of people who take heroin, using the number of arrests for drunken driving to estimate the number of problem drinkers, and using the prevalence of specific antibodies in blood samples to determine the number of victims of a particular disease. In each case, the indicator variable is easier to observe accurately than is the whole population.

In order for the indicator approach to be applied successfully, the following conditions must hold. There must be a strong relationship between the indicator variable and the population size. The nature of that relationship must be known (for example, there might be a hundred heroin users for every known death by overdose and this must be known). The relationship must be constant over time and across areas or the analyst must be aware of the variations.

In summary, the tagging and window methods both require the identification of population members and both place special constraints on the sampling procedure. The indicator method frees one from identifying population members but requires a solid understanding of the relationship between the indicator variable and the population size. In the next chapter, we evaluate the utility of each of these methods for estimating the stock and flow of illegal aliens.

CHAPTER 4

WHY ESTIMATING THE POPULATION OF ILLEGAL ALIENS IS DIFFICULT AT THIS TIME

Although the window, tagging, and indicator methods have been used to estimate the size of various other hidden populations, we find that they cannot be used to provide accurate estimates of the size of the illegal alien population at this time. The apparent diversity among illegal aliens and our lack of knowledge regarding that diversity (that is, population dynamics) prevent analysts from meeting the minimum requirements for generating population estimates from all three methods. Moreover, variations in enforcement policy prevent the use of INS apprehension statistics--the most comprehensive data on illegal aliens--in applications of all three methods.

OBSTACLES TO USING THE TAGGING, WINDOW, AND INDICATOR METHODS

Illegal aliens are commonly thought of as young male Mexican nationals who cross our Southwest border. This stereotype ignores the fact of a great deal of variation in both the demographic and the behavioral characteristics of the entire population of illegal aliens, which includes men and women, young and old, educated and illiterate, and individuals from many different countries. Other important differences among illegal aliens are that some work while others do not, some overstay the terms of their entry documents while others enter without documents, and some come for discrete periods of time while others are intent on permanent residence. This population diversity by itself would make application of the tagging, window, or indicator approaches difficult. The fact that we know so little about how this diversity is distributed, however, makes such application impossible.

The major requirement for successful application of the window method is an understanding of the relationship between window and nonwindow periods. With respect to illegal aliens, for example, an attempt could be made to count all those who cross the Chula Vista sector of our southern border (the space of the window) during the month of August (the time). Let us say that by the end of August a total of 5,000 illegal aliens had been observed. If we could reasonably assume that (1) the flow of illegal aliens at Chula Vista was equivalent to the flow at all seven southern border sectors, (2) one-half of all illegal entries took place at the southern border, and (3) the flow was uniform throughout the year, than we could conclude that the total annual flow is 840,000. This total would be arrived at by multiplying the total number observed at the Chula Vista sector

(5,000) by the total number of southern border sectors (7) by the number of months in the year (12) and then doubling the product (420,000) to account for entries through other borders.

The obvious problem with this example is that we do not have enough information about migration patterns to make any of those assumptions. We do not know what percentage of successful southern border crossings occurs at Chula Vista, how southern border crossings compare with other illegal entries, or how flow varies from month to month. These gaps in our knowledge prevent us from successfully using the window approach to estimate the flow of illegal aliens.

We are hampered in much the same way in trying to use the window approach in estimating the stock of illegal aliens. If, for example, we selected Los Angeles as our window, we would not know how the number observed there related to any number in Houston. If we were to observe illegal aliens in Los Angeles in December, we would not know what adjustments to make for probable seasonal fluctuations. And, if we were to look at places of potential employment, we could not adjust for the number of illegal aliens who are not working or who are working as domestics.

Our lack of knowledge about illegal aliens also prevents us from using the tagging method for deriving accurate estimates of population size. Statistical models for the tagging method require that all population members have equal chances of being tagged or requires us, if the probabilities vary, to know how their chances differ. In practical terms, this means that if different illegal aliens have different chances of being observed, we must know what these differences are if we are going to employ the tagging method.

It seems evident, however, that some illegal aliens are more easily observable than others. Some cross the border without documents while others pass border inspection and later violate the terms of their documents. Some are employed in factory settings, where they may be more visible than others employed as domestics. Recent immigrants may appear to be more foreign (in their speech or living habits, for example) than those who have lived in the United States for several years. In other words, population estimation using the tagging approach is hindered by not knowing exactly how modes of entry, employment patterns, length of stay, and other factors such as health, education, and country of origin affect an individual's chances of being detected.

Using the indicator method is also precluded by current information gaps about illegal aliens. Estimation using the indicator approach requires that one know the magnitude of the relationship between the indicator variable and population size. One example of the use of the indicator approach is Robinson's

(1980) study, in which the number of unaccounted-for deaths (to white males aged 20-44) was used as an indicator of the size of the illegal alien population. The results of that study are questionable because one knows neither the proportion of illegal aliens who are white males aged 20-44 nor the magnitude of the relationship between unaccounted-for deaths and illegal alien population size (that is, the death rate for illegal aliens).

DRAWBACKS IN USING THE INS APPREHENSION DATA

There has been continued interest in whether apprehension data--the most comprehensive data on illegal aliens--can be useful in estimating the size of the illegal alien population. The major obstacle to using INS apprehension data is that we do not know how many illegal aliens are apprehended in relation to how many there are to apprehend. Knowing the value of this "apprehension ratio" would allow us to determine the size of the population. For example, if 100,000 illegal aliens were apprehended in the interior of the country and if we knew that this apprehension ratio was 10 percent, we could conclude that 1,000,000 illegal aliens reside in the United States. Or, if 50,000 illegal aliens were apprehended at the border and if we knew that this apprehension ratio was 20 percent, we could conclude that the annual flow of illegal aliens into the country was 250,000.

Either apprehension ratio depends directly on INS enforcement policy, including elements of that policy like staffing levels and morale, surveillance and transportation equipment, and funding within INS. We could reasonably assume that as any of these change, the apprehension ratio will also change. As long as we do not know exactly how INS changes affect the apprehension ratio--how, for example, introducing a helicopter at a Border Patrol station changes the percentage of illegal border crossers actually apprehended--we cannot use the total number of apprehensions at the border or in the interior to determine the size of either the flow or the stock of illegal aliens.

The utility of INS apprehension data for estimation purposes becomes dubious also when any enforcement unit reaches its maximum apprehension capability. In other words, if 1,000 is the maximum number of apprehensions a given INS unit can make given current resources, then when that unit has apprehended 1,000 illegal aliens any increase in the number of illegal aliens within that unit's area of responsibility will not be reflected in any increase in apprehensions. That is, estimates of the population size from INS apprehension data reflect INS's enforcement capabilities as assuredly as they reflect the actual population.

An additional problem with making population estimates from INS apprehension data is the difficulty in distinguishing between

the number of illegal aliens apprehended and the number of apprehensions. There is considerable evidence that many undocumented aliens, once deported, attempt re-entry shortly thereafter. If they are apprehended again, the INS data will reflect two apprehensions of aliens when, in fact, there is only one alien involved. Using the apprehension data by themselves for estimating the flow or the stock of illegal aliens thus poses considerable difficulty. Notwithstanding, these data play other important roles in a variety of policy decisions.

CHAPTER 5

OTHER APPROACHES FOR MEETING THE NEEDS OF THE CONGRESS FOR INFORMATION ON ILLEGAL ALIENS

Current estimates of the size of the illegal alien population in the United States are unsatisfactory and it seems unlikely that more precise estimates can be derived soon. However, some important information needs of the Congress and the U.S. Immigration and Naturalization Service can be met without precise nationwide estimates of the absolute number of illegal aliens. In this chapter, we present three alternatives to accurately estimating that population that could, nevertheless, meet some of the varied needs of policymakers. They are (1) to use a multi-indicator approach for measuring relative change in population size, (2) to use special studies to estimate the size of important subpopulations, and (3) to expand on current research on immigration with the aim of improving future nationwide population estimates by filling in present information gaps. The multi-indicator approach should provide the most immediately useful information.

THE MULTI-INDICATOR APPROACH FOR MEASURING RELATIVE CHANGE

In formulating policy on illegal aliens, knowing the size of that population is useful for determining whether the number of illegal aliens is so large as to constitute a problem and, if it is, what response is appropriate to direct at that problem. Similarly, in evaluating policy, having accurate information on the number of illegal aliens is also useful. For example, if one does not know the population size either before or after the implementation of a policy, determining the effect of that policy on population size would seem difficult. The key to the multi-indicator approach, however, is that absolute estimates of population size are useful but may not always be necessary for making and evaluating policy. The following example illustrates this point.

Let us assume that we know the actual number of illegal aliens entering this country annually and that it is 500,000. In an effort to decrease this flow, legislation is enacted that makes it more difficult for these individuals to obtain employment. In the year after the legislation's enactment, the actual number of entries drops to 250,000. After eliminating competing explanations for this change, such as an increase in INS funding or a slump in the domestic economy, we conclude that the legislation was successful in cutting the flow by 50 percent.

Now let us suppose that the actual numbers in this example remain the same and that we have no way of knowing them. To

Table 1

The Requirements of Three Estimation Methods

Method	Requirements: Absolute measurement	Requirements: Relative measurement
Window	Relationship between window and nonwindow must be known.	Relationship between window and nonwindow must remain constant.
Tagging	Individuals must be identified. Differential capture probabilities must be known.	Individuals must be identified. Differential capture probabilities must remain constant.
Indicator	Relationship between indicator and population size must be known.	Relationship between indicator and population size must remain constant.

evaluate the legislation, we must estimate the flow of the population for the years before and after the legislation. The method that we use generates an estimate of 1,000,000 entries cut to 500,000. Both numbers would be off by a factor of 2. Nevertheless, the relative change in flow based on the estimated numbers parallels the relative change based on the actual numbers--that is, the 50 percent decrease. Similarly, if we used some other method from which we estimated that the figures for each year were 300,000 and 150,000, respectively, we would have reached the same conclusion about relative change.

The point is that it is not always necessary to know the actual numbers of illegal aliens in order to evaluate policy. When the concern is with change in population size, relative measures are just as useful as absolute measures. Throughout this report, we have indicated that absolute measures of population size--and, consequently, absolute measures of change--cannot currently be estimated accurately. Many of the problems encountered with absolute measures, however, do not exist for relative measures.

The significance of the shift from absolute to relative measurement is that it removes some of the major requirements of the window, tagging, and indicator estimation methods. As we demonstrated in chapter 4, application of these methods requires more information on illegal alien population characteristics than we presently have. However, as table 1 implies, relative measurement requires that the relationships remain constant but not that their magnitude be known. In effect, this

means that we can accurately estimate changes in population size with very little information about population characteristics if we can make the required assumptions of constancy. Bias, for example, can be present in our estimates as long as that bias remains stable over time. If our initial estimates were off by a factor of 3, the percentage change in the population will be accurately reflected as long as all other estimates are off by the same factor.

Although single indicators have been used to determine relative change (by Morris and Mayio, 1980, for example), we suggest that multiple measures be used instead. One benefit of relying on multiple measures derives from the fact that we are never really sure how much bias exists within any single measure. Some measures may lead to exaggerated overestimates of change in population size, whereas others may underestimate it. Using a number of carefully chosen indicators can average out such variations.

It must be cautioned, however, that bias within any single indicator may not be constant--an estimate at time 1 may be off by 10 percent whereas an estimate at time 2 may be off by 20 percent. This is especially problematic if we have no way of knowing whether such change occurred. Therefore, we should exercise care in selecting independent indicators if we are to minimize the effect of potential change in bias. For example, suppose we chose as two independent measures the number of INS apprehensions and the number of unexpected deaths among white males 20-44 years old. The first measure would be influenced by changes in INS enforcement policy; the second measure would be affected by shifts in death rates for illegal aliens. In other words, these two measures would not <u>both</u> be affected by changes in any <u>single</u> factor other than the size of the illegal alien population. This independence may not completely eliminate the effects of change in bias, but it can help minimize them.

We call this approach to measuring relative change the "multi-indicator" approach. The term is derived from the fact that the approach uses a variety of measures as indicators of change. In table 2 on the next page, we illustrate one application of arriving at an estimate of change in population size with this approach.

All five indicators in the table show an increase of 20 to 25 percent in the course of one year. From the data, we can reasonably assume that this percentage accurately reflects the change in the stock of illegal aliens during that year. The same procedure could be used to estimate flow, using indicators such as border apprehensions and applications for temporary visas.

This example of the multi-indicator approach presents an ideal outcome. When the estimates of change converge upon some number, as in the example, the analyst can have considerable

Table 2

A Hypothetical Application
of the Multi-Indicator Approach

Indicator	January 1980	January 1981	Change
1. Applications for change in immigrant status	2,000	2,400	+20%
2. INS interior apprehensions	5,000	6,000	+20%
3. Police contacts with illegal aliens in Los Angeles and Houston	1,500	1,800	+20%
4. Illegal alien population estimate based on unexpected deaths	4,000,000	5,000,000	+25%
5. Population estimate based on CPS-Census match	6,000,000	7,500,000	+25%

confidence in that number as a reliable estimate of change--in this case, of population size. If one or two estimates diverge from the rest, the analyst can study the extenuating circumstances of the time period in question for their possible contribution of bias. If a convincing explanation is found, the analyst can have confidence again in the number upon which the other estimates of change are converging. If the estimates of relative change do not converge, then, of course, the analyst cannot confidently estimate relative change, but in any case the data will be an improvement over what was previously available and may be useful for specific policy questions.

ESTIMATING SUBPOPULATIONS THROUGH TARGETED STUDIES

While our focus has been on estimates of the number of illegal aliens nationwide, many policy questions require estimates of specific components of that population. For example, a policymaker interested in the demand that illegal aliens make on local school systems would want estimates of how many illegal aliens are children of school age. A State government might be concerned only with the population of illegal aliens in that State. If INS seeks to develop a policy with regard to people who violate the terms of their visas, the population of concern would not include undocumented aliens.

The number of subpopulations of interest is almost as limitless as the number of policy issues related to immigration. Moreover, the current inability to estimate the total population size accurately does not necessarily reflect an inability to estimate subpopulations. For example, as we indicated in chapter 2, the only serious obstacles to estimating the number of people who overstay their visas are practical problems with collecting and processing records. A detailed listing of which components of the illegal alien population can and cannot be estimated accurately is beyond the scope of this report. Making this determination depends on a close examination of the specific populations of interest, their assumed diversity, the available methods of identification, and the appropriateness of applying available estimation methods. In general, the ease with which one can estimate depends on the stability, homogeneity, and identifiability of members of the subpopulation.

We do not, however, recommend combining the results of individual estimates of subpopulations to produce one national estimate of the total population of illegal aliens in the country. Not only is it not feasible to conduct a sufficiently large number of special studies to include all illegal aliens, but also the subpopulations would undoubtedly overlap, leading to the counting of some people more than once. We mentioned an example of this in chapter 4 when discussing some of the drawbacks in using INS apprehension data--an illegally entering migrant who is stopped at the border but tries another form of entry the next

day or within that same year may or may not be detected and counted as a second "apprehension."

FURTHER RESEARCH TO FILL THE INFORMATION GAPS

The population of illegal aliens in the United States is highly diverse. Full information on this diversity is lacking--information that is necessary if estimation methods are to be applied successfully. To help fill in the information gaps, conclusive findings on almost any aspect of illegal immigration would be helpful. Ethnographic research involving intensive observation of individuals and small groups would expand our understanding of the varieties of behavior among illegal aliens and would help us apply that understanding in attempting to estimate population size. Surveys would help reveal how these varieties of behavior are distributed throughout the population at large. Experiments similar to studies INS has carried out previously would help determine the effects of change in INS's staffinf, resources, and strategies on its apprehension rates.

A combination of ethnographic research, surveys, and experiments would require extensive resources and substantial amounts of time for both planning and execution. The merits of implementing a comprehensive research program of this kind would depend on the extent of congressional concern for reliable, narrow estimates weighed against the significant expenditure of resources that would be required at both the national and the local levels.

AGENCY COMMENTS AND OUR RESPONSE

The U.S. Department of Justice reviewed a draft of this report and has concurred with our major findings. This final version of the report incorporates changes that Justice suggested, and we reprint the agency's comments together with our response in appendix V.

APPENDIX I

United States Senate

COMMITTEE ON THE JUDICIARY
WASHINGTON, D.C. 20510

STROM THURMOND, S.C., CHAIRMAN
CHARLES McC. MATHIAS, JR., MD.
PAUL LAXALT, NEV.
ORRIN G. HATCH, UTAH
ROBERT DOLE, KANS.
ALAN K. SIMPSON, WYO.
JOHN P. EAST, N.C.
CHARLES E. GRASSLEY, IOWA
JEREMIAH DENTON, ALA.
ARLEN SPECTER, PA.

JOSEPH R. BIDEN, JR., DEL.
EDWARD M. KENNEDY, MASS.
ROBERT C. BYRD, W. VA.
HOWARD M. METZENBAUM, OHIO
DENNIS DeCONCINI, ARIZ.
PATRICK J. LEAHY, VT.
MAX BAUCUS, MONT.
HOWELL HEFLIN, ALA.

EMORY SNEEDEN, CHIEF COUNSEL
QUENTIN CROMMELIN, JR., STAFF DIRECTOR

SUBCOMMITTEE ON
IMMIGRATION AND REFUGEE POLICY
ALAN K. SIMPSON, WYO., CHAIRMAN
STROM THURMOND, S.C.
CHARLES E. GRASSLEY, IOWA
EDWARD M. KENNEDY, MASS.
DENNIS DeCONCINI, ARIZ.

RICHARD W. DAY, CHIEF COUNSEL AND STAFF DIRECTOR

October 26, 1981

Mr. Charles Bowsher
Comptroller General
General Accounting Office
441 G. Street N. W.
Washington, D. C. 20548

Dear Mr. Bowsher:

The Senate Subcommittee on Immigration and Refugee Policy is currently reviewing a variety of policy options for dealing with the problem of illegal aliens. In the course of our work it has become obvious that accurate and reliable estimates of the size of the illegal alien population are important. Initial discussions between Arnold Leibowitz of my staff and staff from your Institute for Program Evaluation indicate that your organization may be able to provide us with some assistance in making such estimates.

Specifically, the Subcommittee is interested in having a review and analysis done of methodologies which have been developed and currently used for measuring other "hidden populations" (e.g., drug users, tax evaders, cancer victims, etc.) in order to see what potential these procedures may have for improving estimates in the area of illegal aliens.

It would be most helpful if GAO staff could brief the Subcommittee staff on their findings by March 1, 1982 and then prepare a report on their findings for later submission.

Most sincerely,

Alan K. Simpson
Chairman

AKS:als

APPENDIX II

ILLEGAL ALIEN POPULATION ESTIMATES

We have adapted this list of estimates from the appendix in our 1981 report on undocumented aliens (GAO, 1981). As in the text, the abbreviated bibliographic citations are expanded in full in appendix IV. Items marked with an asterisk (*) are reviewed in our discussion in chapter 2.

Estimate in millions	Year of estimate	Source
1.6 = Mexican-origin	1970	Goldberg, 1974*
1.0	1972	INS Commissioner R. Farrell, appropriations hearings, 1972
3.9 = 2.9-5.7 18-44 years	1973	Lancaster and Scheuren, 1978*
4.0-12.0	1975	INS Commissioner L. F. Chapman, 1975
8.2 = 4.2-11.0 5.2 = Mexican-origin 0.97 = flow from Mexico	1975	Lesko Associates, 1975*
0.6-4.7 = white males 20-44 years	1975	Robinson, 1980*
0.4-1.2 = increase in Mexican-origin population 1970-1975 0.08-0.2 = flow from Mexico	1975	Heer, 1979*
0.5 = flow of fraudulent entries	1975	U.S. Department of Justice, 1976*
6.0	1976	INS Commissioner L. F. Chapman, statement before Subcommittee on Immigration and Naturalization, Committee on the Judiciary, U.S. Senate, 94th Cong., 2nd sess., Washington, D.C., March 17, 1976.
1.7 = overstays	1976	Chapman, 1976*
0.18-0.38 = flow of overstays by air	1974-77	Vining, 1979*

APPENDIX II

Estimate in millions	Year of estimate	Source
0.5-1.2 = Mexican-origin	1977	CENIET, 1979a*
0.7-2.2 = Mexican-origin	1977	U.S. Bureau of the Census, re-estimation of the Mexican Border Survey, 1979
4.3-6.2 2.4 = Mexican-origin 1.05 = flow from Mexico	1977	Morris and Mayio, 1980*
1.075-1.735 = flow from Mexico	1978	Morris and Mayio, 1980*
3.0-6.0	1978	INS Commissioner L. Castillo, statement before Select Committee on Population, U.S. House of Representatives, 95th Cong., 2nd sess., Washington, D.C., April 6, 1978
1.1-4.1 = 20-44 years	1978	J. G. Robinson, "Updating of Estimates of Illegal Aliens Based on Analysis of Trends in Age-Specific Death Rates," 1981, in Warren, 1981
3.5-5.0 1.5-2.5 = Mexican-origin	1978	Siegel, Passel, and Robinson, 1981
0.4 = Mexican-origin 15+ years	1978-79 Dec.-Jan.	CENIET, 1979b*
2-12 = although emerging consensus seems to be 3.0-6.0	1979	Select Committee on Population Report, U.S. House of Representatives, 1979
1.025-1.475 = illegal aliens included in CPS	1979	Warren, 1981*
3.0-6.0	1981	U.S. Select Commission on Immigration and Refugee Policy, 1981

APPENDIX III

GLOSSARY

Bias. Systematic error.

Biased sample. A set of observations on a sample that do not accurately reflect the nature of that population.

Demographic characteristic. An ascribed or achieved status characteristic such as age, gender, marital status, nationality, and race.

Deportable alien. An alien who has violated immigration law in such fashion as to be liable to deportation whether apprehended or not.

Differential capture probabilities. Unequal chances of being observed.

Discrepancy model. Observing the difference between two independent measures of the same population.

Ethnographic research. Intensive observation of individuals or small groups and their social environment.

Flow. The number of illegal aliens entering the country within a period of time.

- Gross flow. The total number entering.
- Net flow. The number entering minus the number exiting; that is, the change in population size.

Fraudulent entry. An inspected entry into the country with either false or stolen documents or with the intention of violating the terms of proper documents.

Got-away ratio. The number of successful illegal entries into the country divided by the number of unsuccessful (that is, apprehended) entries into the country.

Hidden population. Any population whose members are either difficult to observe or, upon observation, difficult to identify as members of that population.

Independent measures. Data collection mechanisms not jointly affected by any factors other than the object to be measured.

Indicator method. Observing some variable that is known to be strongly related to population size rather than observing the population size directly.

INS apprehension data (or statistics). INS enumerations of the number of apprehensions it made as a result of violations of any of a number of immigration laws.

Multiple entries. More than one entry or attempted entry by any given alien in a specified period of time.

Multiple-indicator method. Observing several independent measures known to be related to population size to achieve consensus on an estimate.

Overstay. Remaining in the country as an alien beyond the expiration of a temporary visa.

Permanent resident alien. An alien who has secured the legal right to live and work in the United States for an indeterminate time.

Relative change. Change in size over time expressed as a proportion of initial size.

Reliable. Stable or consistent with repetition over time.

Reverse sampling. Estimating the size of a population from the number of observations in a sample.

Stock. The number of illegal aliens residing in the country at a particular moment; the resident population.

Sub rosa. Secret; confidential; hidden.

Tagging method. Estimating the size of a population from the frequency with which members of that population are encountered.

Underenumeration. A counting of a population that fails to include all the members of that population; undercoverage.

Undocumented alien. An alien who is without immigration papers and who has probably entered without inspection and not at an official INS point of entry.

Valid. Appropriate; true; correct.

Visa abuser. An alien who violates the terms of a properly obtained immigration visa, as when failing to leave upon its expiration or taking unauthorized employment.

Window method. Counting all members of a population observed during a specific period of time and within a specific geographic area.

APPENDIX IV

SELECTED BIBLIOGRAPHY

Baca, R., and D. Bryan. "Mexican Undocumented Workers in the Binational Community: A Research Note." International Migration Review, 15 (Winter 1981), 737-48.

Burnham, K. P., and W. S. Overton. "Estimation of the Size of a Closed Population When Capture Probabilities Vary Among Animals." Biometrika, 65 (1978), 625-33.

CENIET (Centro Nacional de Informacion y Estadisticas del Trabajo). El Volumen de la Migracion de Mexicanos No Documentados a los Estados Unidos: Nuevas Hipotesis, by Manuel Garcia y Griego. Mexico: December 1979a.

-----. Los Trabajadores Mexicanos en los Estados Unidos: Primeros Resultados de la Encuesta Nacional de Emigracion, by Carlos H. Zazueta and Rodolfo Corona. Mexico: December 1979b.

-----. La Encuesta Nacional de Emigracion a la Frontera del Pais y a los Estados Unidos: Descripcion del Proyecto y Hallazgos de la Segunda Etapa, by Geronimo Martinez Garcia. Mexico: August 1978.

Chapman, L. F. Address before the Michigan Associated Press Editorial Association, Gaylord, Michigan, June 11, 1976.

Community Research Associates, Inc. Undocumented Immigrants: Their Impact on the County of San Diego. San Diego, Calif.: May 1980.

Cornelius, W. A. Illegal Migration to the United States: Recent Research Findings, Policy Implications and Research Priorities. Cambridge, Mass.: M.I.T., Migration and Development Study Group, May 1977.

DOJ (U.S. Department of Justice, Immigration and Naturalization Service). Fraudulent Entrants Study: A Study of Malafide Applicants for Admissions at Selected Airports and Southwest Land Border Parts, Illegal Alien Study--Part I. Washington, D.C.: September 1976.

Dondero, L. G., G. R. Bieber, and R. E. Forrester. Final Report, Evaluation of Capability of the Helicopter in U.S. Border Patrol Operations. Vol. 2. Analysis of Flow of EWIs. Staffing Experiment and Sensor Reporting Systems. McLean, Va.: General Research Corp., March 1977.

Furst, C. J., L. J. Beckman, and C. Y. Nakamura. "Validity of Synthetic Estimates of Problem-Drinker Prevalence." American Journal of Public Health, 71 (September 1981), 1016-20.

GAO (U.S. General Accounting Office). Number of Undocumented Aliens Residing in the United States Unknown, GGD-81-56. Washington, D.C.: April 6, 1981.

-----. Illegal Aliens: Estimating Their Impact on the United States, PAD-80-22. Washington, D.C.: March 14, 1980a.

-----. Prospects Dim for Effectively Enforcing Immigration Laws, GGD-81-4. Washington, D.C.: November 5, 1980b.

Goldberg, H. "Estimates of Emigration from Mexico and Illegal Entry into the United States, 1960-1970, by the Residual Method." Graduate research paper, Georgetown University, Center for Population Research, Washington, D.C., 1974.

Greene, M. A., and S. Stollmack. Estimating the Number of Criminals. Washington, D.C.: ANALOGS, Inc., for U.S. Department of Justice, National Institute of Justice, 1980.

Heer, D. M. "What Is the Annual Net Flow of Undocumented Mexican Immigrants to the United States?" Demography, 16 (August 1979), 417-23.

Keely, C. B. "Illegal Migration." Scientific American, 246 (March 1982), 41-47.

Lancaster, C., and F. J. Scheuren. "Counting the Uncountable Illegals: Some Initial Statistical Speculations Employing Capture-Recapture Techniques." 1977 Proceedings of the Social Statistics Section, Part I. American Statistical Association, 1978.

Leatherwood, S. "Evaluation of Some Techniques for Aerial Census of Bottlenosed Dolphins." Journal of Wildlife Management, 42 (April 1978), 239-50.

Lesko Associates. Final Report: Basic Data and Guidance Required to Implement a Major Illegal Alien Study During Fiscal Year 1976. Washington, D.C.: U.S. Department of Justice, Immigration and Naturalization Service, October 1975.

Morris, M. D., and A. Mayio. Illegal Immigration and United States Foreign Policy. Washington, D.C.: The Brookings Institution for U.S. Department of Labor, October 1980.

North, D. S. Illegal Alien Study Design. Vol. 1. Final Report. Washington, D.C.: Linton & Co. for U.S. Department of Justice, Immigration and Naturalization Service, May 1975.

Person, P. H., R. L. Retka, and J. A. Woodward. A Method for Estimating Heroin Use Prevalence. Washington, D.C.: National Institutes of Health, National Institute of Drug Abuse, 1977.

Piore, M. J. "The Economic Role of Migrants in the U.S. Labor Market." Sourcebook on the New Immigration: Implications for the United States and the International Community, ed. by R. S. Bryce-Laporte. New Brunswick, N.J.: Transaction Books, 1980.

Reichert, J., and D. S. Massey. "History and Trends in U.S. Bound Migration from a Mexican Town." International Migration Review, 14 (Winter 1980), 475-91.

-----. "Patterns of U.S. Migration from a Mexican Sending Community: A Comparison of Legal and Illegal Migrants." International Migration Review, 13 (Winter 1979), 599-623.

Robinson, J. G. "Estimating the Approximate Size of the Illegal Alien Population in the United States by the Comparative Trend Analysis of Age-Specific Death Rates." Demography, 17 (May 1980), 159-76.

Siegel, J. S., J. S. Passel, and J. G. Robinson. "Preliminary Review of Existing Studies of the Number of Illegal Residents in the United States [January 1980]." U.S. Select Commission on Immigration and Refugee Policy, U.S. Immigration Policy and the National Interest: Supplement to the Final Report and Recommendations. Appendix E. Policy Papers on Illegal Migration to the United States. Washington, D.C.: April 30, 1981.

Silberman, G. "Recent Advances in Evaluation Methods." Handbook of Criminal Justice Evaluation, ed. by M. Klein and K. Teilman. Beverly Hills, Calif.: Sage Publications, 1980.

Ugalde, A., F. D. Bean, and G. Cardenas. "International Migration from the Dominican Republic: Findings from a National Survey." International Migration Review, 13 (Summer 1979), 235-54.

U.S. Select Commission on Immigration and Refugee Policy. U.S. Immigration Policy and the National Interest: Supplement to the Final Report and Recommendations. Washington, D.C.: April 30, 1981.

Valentine, P. W. "Aliens No Longer Have to Report in Every Year." The Washington Post, January 12, 1982, p. A15.

Vining, D. R. Net Migration by Air: A Lower-Bound on Total Migration to the United States. Working Papers in Regional Science and Transportation No. 15. Philadelphia, Pa.: University of Pennsylvania, August 1979.

Warren, R. "Estimation of the Size of the Illegal Alien Population in the United States." Paper presented at the Census Advisory Committee meeting of the American Statistical Association, Washington, D.C., November 1981.

APPENDIX V

AGENCY COMMENTS AND OUR RESPONSE

The letter from the Department of Justice commenting on a draft of this report begins below and continues on page 34. The note numbers we have added at the lefthand margins are keyed to our paragraph-by-paragraph response on page 35.

U.S. Department of Justice

Washington, D.C. 20530

SEP 3 1982

Mr. William J. Anderson
Director
General Government Division
United States General Accounting Office
Washington, D.C. 20548

Dear Mr. Anderson:

This letter is in response to your request to the Attorney General for the comments of the Department of Justice (Department) on your draft report entitled "Problems and Options in Estimating the size of the Illegal Alien Population."

The Department has reviewed the subject report and found it to be very informative. Overall, we agree with the General Accounting Office's (GAO) analyses of the problems associated with reliably estimating the size of the illegal alien population, and we concur that policymakers need accurate estimates of the number of illegal aliens in the country as well as the number entering each year.

First, we would like to express our appreciation to GAO for their display of perseverance and technical knowledge in analyzing previous estimation attempts to assess the size of the illegal alien population and in proposing other approaches for supplying the information needs to both the Congress and the Immigration and Naturalization Service (INS). We support GAO's conclusion that the multi-indicator approach should provide the most immediately useful information.

Based on our review of the report, we are offering several comments on GAO's analyses of previous studies cited in the report and on suggested changes to a footnote and to definitions used in the glossary. We believe our comments, which follow, will be helpful in increasing the accuracy and completeness of the report.

Note 1--The description of the Lesko study on page II-6 does not state or explain part of the methodology used. In addition to INS apprehension data discussed in the report, which were used to estimate the Mexican component of the illegal alien population, Lesko Associates used a modified "delphi" technique to estimate the size of the total illegal alien population. We believe this second part of their methodology, which was the basis of their 8.2 million estimate, should be included in the GAO analysis.

Note 2 --Although not an empirical methodology in the true sense, the Chapman estimates mentioned on page II-3 were based on polling INS district directors who presumably had some idea of the magnitude of the size of the problem within their jurisdictions. This "methodology" might be worth mentioning in the GAO report, as appropriate, for completeness.

Note 3 --The 740,000 unmatched documents referenced on page II-3 relate to a direct by-product of the INS nonimmigrant document control system--arrival documents in the system to which departure documents were not matched--rather than to a sample taken for the purpose of estimating overstays. These numbers were upper bounds of overstays, since INS investigative efforts on samples of such records tended to show that some such persons had, in fact, departed the country or were still present but in a legal status. Minor revision of this section of the report could be made to make it more accurate.

Note 4 --GAO's criticism of the Fraudulent Entrants Study on page II-7 may not be completely valid. Based on information provided by Mr. David North of the New TranCentury Foundation, who designed and analyzed the results of the Fraudulent Entrants Study under contract to INS, we understand that the sampling and procedural techniques used in the Fraudulent Entrants Study were statistically valid. Unfortunately, the criticisms made about this study, such as those by Dr. Charles Keely in his article cited in Appendix I of the GAO report, were based on a summary of the study rather than the full report which documented fully the procedures and sampling techniques used. It appears, based on Appendix I again, that GAO used the summary rather than the full report. This is an area where additional discussion, if feasible, might be warranted to ensure that GAO's report is as accurate as possible in its analysis of previous studies.

Note 5 --Footnote 2, page V-2, reads "successfully controlled for." We wonder if it should read "statistically controlled for." Either wording could be correct depending upon the author's intent.

Note 3 --The definition of "INS apprehension data (or statistics)" in the glossary should be corrected. As correctly stated on pages IV-4 through IV-6, those data pertain to the number of apprehensions, not the number of people. Multiple apprehensions of the same person are possible.

Note 3 --The "Got-away ratio" in the glossary should read "Number of successful illegal entries . . . ."

We appreciate the opportunity to comment on the report. Should you desire any additional information, please feel free to contact me.

Sincerely,

Kevin D. Rooney

Kevin D. Rooney
Assistant Attorney General
for Administration

Note 1: To our description of Lesko Associates' flow estimate, we have added a footnote describing this method of estimating stock. See page 9.

Note 2: As a "methodology," combining educated guesses seemed to be adequately addressed in the explanation of the Delphi method. Additionally, the estimate on page II-3 (now page 6) was derived (as we have described in the report) from matching arrival and departure forms.

Note 3: We have made the revision suggested.

Note 4: We reviewed North's original report and Keely's critique of the study. While agreeing that the within-site sampling appears to have been valid, we have retained our additional criticism that the resulting estimates would not be stable over time.

Note 5: We deleted the footnote.

(973169)

Request for copies of GAO reports should be sent to:

U.S. General Accounting Office
Document Handling and Information
Services Facility
P.O. Box 6015
Gaithersburg, Md. 20760

Telephone (202) 275-6241

The first five copies of individual reports are free of charge. Additional copies of bound audit reports are $3.25 each. Additional copies of unbound report (i.e., letter reports) and most other publications are $1.00 each. There will be a 25% discount on all orders for 100 or more copies mailed to a single address. Sales orders must be prepaid on a cash, check, or money order basis. Check should be made out to the "Superintendent of Documents".