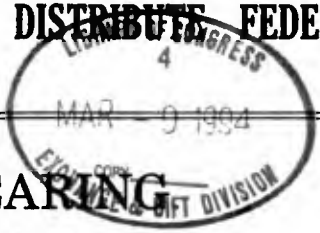
United States 111

JOINT HEARING ON H.R. 6 AND H.R. 1645: USES OF CENSUS DATA TO DISTRIBUTE FEDERAL FUNDING



JOINT HEARING

BEFORE THE

SUBCOMMITTEE ON ELEMENTARY, SECONDARY, AND VOCATIONAL EDUCATION

OF THE

COMMITTEE ON EDUCATION AND LABOR

AND THE

SUBCOMMITTEE ON CENSUS, STATISTICS, AND POSTAL PERSONNEL

OF THE

COMMITTEE ON POST OFFICE AND CIVIL SERVICE

HOUSE OF REPRESENTATIVES

ONE HUNDRED THIRD CONGRESS

FIRST SESSION

JOINT HEARING HELD IN WASHINGTON, DC, JULY 13, 1993

Education and Labor Com. Serial No. 103–37
Post Office and Civil Service Com. Serial No. 103–17

Printed for the use of the Committee on Education and Labor and the Committee on Post Office and Civil Service

U.S. GOVERNMENT PRINTING OFFICE

75–540 ≈ WASHINGTON : 1994

For sale by the U.S. Government Printing Office
Superintendent of Documents, Congressional Sales Office, Washington, DC 20402
ISBN 0-16-043519-6

COMMITTEE ON EDUCATION AND LABOR

WILLIAM D. FORD, Michigan, *Chairman*

WILLIAM (BILL) CLAY, Missouri
GEORGE MILLER, California
AUSTIN J. MURPHY, Pennsylvania
DALE E. KILDEE, Michigan
PAT WILLIAMS, Montana
MATTHEW G. MARTINEZ, California
MAJOR R. OWENS, New York
THOMAS C. SAWYER, Ohio
DONALD M. PAYNE, New Jersey
JOLENE UNSOELD, Washington
PATSY T. MINK, Hawaii
ROBERT E. ANDREWS, New Jersey
JACK REED, Rhode Island
TIM ROEMER, Indiana
ELIOT L. ENGEL, New York
XAVIER BECERRA, California
ROBERT C. SCOTT, Virginia
GENE GREEN, Texas
LYNN C. WOOLSEY, California
CARLOS A. ROMERO-BARCELÓ, Puerto Rico
RON KLINK, Pennsylvania
KARAN ENGLISH, Arizona
TED STRICKLAND, Ohio
RON DE LUGO, Virgin Islands
ENI F. H. FALEOMAVAEGA, American Samoa
SCOTTY BAESLER, Kentucky
ROBERT A. UNDERWOOD, Guam

WILLIAM F. GOODLING, Pennsylvania
THOMAS E. PETRI, Wisconsin
MARGE ROUKEMA, New Jersey
STEVE GUNDERSON, Wisconsin
RICHARD K. ARMEY, Texas
HARRIS W. FAWELL, Illinois
CASS BALLENGER, North Carolina
SUSAN MOLINARI, New York
BILL BARRETT, Nebraska
JOHN A. BOEHNER, Ohio
RANDY "DUKE" CUNNINGHAM, California
PETER HOEKSTRA, Michigan
HOWARD P. "BUCK" McKEON, California
DAN MILLER, Florida

PATRICIA F. RISSLER, *Staff Director*
JAY EAGEN, *Minority Staff Director*

SUBCOMMITTEE ON ELEMENTARY, SECONDARY, AND VOCATIONAL EDUCATION

DALE E. KILDEE, Michigan, *Chairman*

GEORGE MILLER, California
THOMAS C. SAWYER, Ohio
MAJOR R. OWENS, New York
JOLENE UNSOELD, Washington
JACK REED, Rhode Island
TIM ROEMER, Indiana
PATSY T. MINK, Hawaii
ELIOT L. ENGEL, New York
XAVIER BECERRA, California
GENE GREEN, Texas
LYNN C. WOOLSEY, California
KARAN ENGLISH, Arizona
TED STRICKLAND, Ohio
DONALD M. PAYNE, New Jersey
CARLOS A. ROMERO-BARCELÓ, Puerto Rico

WILLIAM F. GOODLING, Pennsylvania
STEVE GUNDERSON, Wisconsin
HOWARD "BUCK" McKEON, California
THOMAS E. PETRI, Wisconsin
SUSAN MOLINARI, New York
RANDY "DUKE" CUNNINGHAM, California
DAN MILLER, Florida
MARGE ROUKEMA, New Jersey
JOHN A. BOEHNER, Ohio

COMMITTEE ON POST OFFICE AND CIVIL SERVICE

WILLIAM L. CLAY, Missouri, *Chairman*

PATRICIA SCHROEDER, Colorado
FRANK McCLOSKEY, Indiana
GARY L. ACKERMAN, New York
THOMAS C. SAWYER, Ohio
PAUL E. KANJORSKI, Pennsylvania
ELEANOR HOLMES NORTON, District of Columbia
BARBARA-ROSE COLLINS, Michigan
LESLIE L. BYRNE, Virginia
MELVIN L. WATT, North Carolina
ALBERT RUSSELL WYNN, Maryland
GREG LAUGHLIN, Texas
SANFORD D. BISHOP, Jr., Georgia
SHERROD BROWN, Ohio
ALCEE L. HASTINGS, Florida

JOHN T. MYERS, Indiana
BENJAMIN A. GILMAN, New York
DON YOUNG, Alaska
DAN BURTON, Indiana
CONSTANCE A. MORELLA, Maryland
THOMAS J. RIDGE, Pennsylvania
THOMAS E. PETRI, Wisconsin
SHERWOOD L. BOEHLERT, New York
JIM SAXTON, New Jersey

GAIL E. WEISS, *Staff Director*
ROBERT E. LOCKHART, *General Counsel*
DORIS MOORE-GLENN, *Deputy Staff Director*
JOSEPH A. FISHER, *Minority Staff Director*

SUBCOMMITTEE ON CENSUS, STATISTICS AND POSTAL PERSONNEL

THOMAS C. SAWYER, Ohio, *Chairman*

FRANK McCLOSKEY, Indiana
ALBERT RUSSELL WYNN, Maryland

THOMAS E. PETRI, Wisconsin
THOMAS J. RIDGE, Pennsylvania

TERRIANN LOWENTHAL, *Subcommittee Staff Director*

CONTENTS

JOINT HEARING ON H.R. 6 AND H.R. 1645: USES OF CENSUS DATA TO DISTRIBUTE FEDERAL FUNDING

TUESDAY, JULY 13, 1993

HOUSE OF REPRESENTATIVES,
SUBCOMMITTEE ON ELEMENTARY,
SECONDARY, AND VOCATIONAL EDUCATION,
COMMITTEE ON EDUCATION AND LABOR, AND
SUBCOMMITTEE ON CENSUS, STATISTICS,
AND POSTAL PERSONNEL,
COMMITTEE ON POST OFFICE AND CIVIL SERVICE,
Washington, DC.

The subcommittees met, pursuant to call, at 10:05 a.m., Room 2175, Rayburn House Office Building, Hon. Dale E. Kildee, presiding.

Members present: Representatives Kildee, Sawyer, Unsoeld, Reed, Roemer, Becerra, Green, Woolsey, Payne, Gunderson, and McKeon.

Staff present: Subcommittee on Elementary, Secondary, and Vocational Education: Susan A. Wilhelm, staff director; S. Jefferson McFarland, III, legislative counsel; Jack Jennings, general counsel, full committee; Diane Stark, legislative specialist, full committee; and Jane Baird, minority counsel; Subcommittee on Census, Statistics, and Postal Personnel: TerriAnn Lowenthal, staff director; Chris Collins, staff assistant; and George Conant, minority staff assistant.

Chairman KILDEE. The Subcommittee on Elementary, Secondary, and Vocational Education convenes this morning for a joint hearing with the Subcommittee on Census, Statistics, and Postal Personnel. I am pleased to be chairing this hearing jointly with Chairman Sawyer—we can cut the gavel in half, if you wish. His work as the Chairman of the census subcommittee is very well known. It is an added bonus that Mr. Sawyer is Chairman of the census subcommittee and a valued member of this subcommittee as well.

Our topic this morning is the availability and quality of census data used in Federal funding formulas. Chairman Sawyer has introduced legislation that would direct the Census Bureau to biannually update the count of poor children aged 5 through 17. Clearly, such data would go a long way in the distribution of Chapter 1 funds which currently occurs every 10 years as new census data becomes available. We will also discuss efforts to obtain data on the number of poor children by school districts. As members are aware, current law provides that the Chapter 1 basic grant formula is

used to count the number of poor children by school districts if that data is available. In the absence of such data, county poverty counts are used.

Previous attempts to obtain district level numbers have not resulted in data that could be used to make allocations at the national formula. The formula dates back to a woman by the name of Mrs. Orshansky, I believe. When I first came to Washington 17 years ago, the Orshansky formula—that term was used a great deal—was part of the lore of the distribution of these funds.

Which department did she work for?

Ms. RUGGLES. Social Security.

Mr. ELLIOTT. HEW, Social Security, and now HHS.

Chairman KILDEE. Okay—and now HHS.

At any rate, whenever we talk about Chapter 1 or distribution, we generally refer to Mrs. Orshansky on this.

At this point, I would like to recognize Chairman Sawyer for his opening statement.

Mr. SAWYER. Thank you, Mr. Chairman.

I worry any time you say "at any rate." That has been part of the problem.

I am particularly pleased to join with you this morning. The work that we have done together since I have been in Congress has been perhaps as instructive as any that I have had the opportunity to do since I have been here. The nexus of the issues that bring us here today really goes to the heart of how well we provide assistance not only in education but a number of other arenas.

Today we are looking at the question, as you said, of the availability of data on poverty both from the census every 10 years and in between. I am hopeful that as a result of the extraordinary testimony that we will hear today, we will come to better understand how those estimates are produced, the quality of that data, and what potential we have for producing estimates that are more timely in the future.

I want to emphasize that this isn't just an academic exercise. Poverty data are used to distribute some $20 billion annually in Federal funding and drive distributions within States of dollar amounts that go well beyond that. The interest of this committee is, I think, indicative of the broad policy consequences of the full range of Federal statistical programs.

Right now, the only source of reliable poverty data for small places is the decennial census. It is the clear perception of large numbers of members of this committee and others throughout the country that the infrequent production of that data undermines the ability of many Federal programs to reach their target populations effectively.

In the school year that just ended this past spring, the Chapter 1 grant program was funded using poverty data that reflected 1979 income data. Those numbers are 14 years out of date. If we tried to fund the defense budget based on data that was as removed from current situations as that is, we would simply be laughed off of the world stage.

Chairman KILDEE. There are some, however, who would still try to fund the Defense Department as it was during the height of the Cold War.

Mr. SAWYER. I don't even want to get into that, Mr. Chairman.

When we talk about these data, they aren't just old, they are ancient, they are ancient history.

The change that we are going through right now is profound. We are talking about data that may look precise by virtue of its geographic detail. But it is not just enough for the data to look precise. If the numbers aren't timely, then to the degree that they fail to capture change, they simply aren't accurate.

Change is a dominant characteristic of our time. If the census is occurring only once a decade, we are failing to capture what is one of the most important phenomena of our time, and that is change itself. That is, as much as anything, the reason I introduced the measure that we are reviewing today. It would require the Bureau to produce and publish estimates down to the school district level every two years. It would require separate tabulations of school-age children.

Contrary to some of the things that have been suggested, this is not a simple task. It clearly is going to require further research and refinement of technique. Perhaps even an independent review of selected methodology would be helpful. But it is also clear that more frequent measurements are long overdue. If we can do them and do them affordably, then we ought to do them. There is just too much at stake with the policy that they drive. It would not only help the Congress target the dollars that we appropriate, but it would help us to understand and measure the programmatic success or failure of those programs that are driven by that kind of data.

In any case, I am enormously grateful, Mr. Chairman, not only to you and to Mr. Goodling for having this hearing, but to the Subcommittee on Commerce, Justice, State, and Judiciary for including $600,000 in their 1994 funding bill to begin research and development on an intercensal poverty estimates program.

So, again, Mr. Chairman, thank you, and I look forward to our testimony today.

Chairman KILDEE. And thank you, Mr. Chairman.

When I was going to school, there was emblazoned on the wall the statement, "Knowledge is power," and knowledge should lead us to fairness in this area too, so I think the more knowledge we have, the more fair we can be to the students out there for whom this, particularly Chapter 1, was designed. So no one should be afraid of the truth. We should be seeking the truth so we can be fair.

Mr. SAWYER. Mr. Chairman, I have attached to my scale in my bathroom the saying, "What you measure is what you get."

Chairman KILDEE. Mr. McKeon.

Mr. MCKEON. It is hard to top that.

Thank you, Mr. Chairman. I would like to thank you and Chairman Sawyer for providing the opportunity today to consider H.R. 1645, the Poverty Data Improvement Act of 1993. I would further like to commend Chairman Sawyer for his work on this bill. I am proud to be a cosponsor of this important legislation.

Yearly, billions of Federal dollars are allocated based on poverty data. By updating the poverty data every two years, as this bill requires, we will be most efficiently and effectively using these Fed-

eral dollars for their intended purpose. Currently, poverty figures are taken from the decennial census.

My State of California is one of many in which the population changed drastically during the past 10 years. For example, the number of school children from low-income families increased 38.2 percent, or by 247,162 children, between 1980 and 1990. Since 1990 census data will not be factored into the Chapter 1 formula until the 1993/1994 school year, past year allocations were based on 12-year-old data. This translated in a loss of over $126 million for California in the 1992/1993 school year. This amount of funding can make a big difference to schools and in the individual lives of children who benefit the most from these dollars.

This legislation will ensure not only *updated* but *timely* poverty data as well. To achieve the goals we have in mind when we distribute these dollars, we must use accurate poverty figures. I urge my colleagues to listen closely to the testimony today and to give this legislation serious consideration.

Thank you, Mr. Chairman.

Chairman KILDEE. Thank you, Mr. McKeon.

Our witnesses this morning are Mr. William Butz, associate director for demographic programs, the Bureau of the Census, U.S. Department of Commerce; the Honorable Emerson Elliott, commissioner of education statistics, U.S. Department of Education; and Ms. Patricia Ruggles, senior research associate, the Urban Institute, Washington, DC.

Mr. Butz, you may begin.

STATEMENTS OF WILLIAM P. BUTZ, ASSOCIATE DIRECTOR FOR DEMOGRAPHIC PROGRAMS, BUREAU OF THE CENSUS, U.S. DEPARTMENT OF COMMERCE, WASHINGTON, DC; HON. EMERSON J. ELLIOTT, COMMISSION OF EDUCATION STATISTICS, U.S. DEPARTMENT OF EDUCATION, WASHINGTON, DC; AND PATRICIA RUGGLES, SENIOR RESEARCH ASSOCIATE, THE URBAN INSTITUTE, WASHINGTON, DC

Mr. BUTZ. Thank you, Mr. Chairman; and thank you also, Mr. Chairman. It is really a great pleasure for me to be here on behalf of the Census Bureau, and I am honored to be here personally not only appearing before this joint meeting of the two subcommittees but also having the opportunity to appear with two very highly respected and esteemed colleagues.

I have submitted written testimony that I would appreciate being put in the record.

Chairman KILDEE. Yes. Without objection, that will be included in its entirety.

Mr. BUTZ. Thank you, sir, and I will summarize it rather quickly and then be very happy to try to answer any questions you may have.

In the first part of these remarks, I will provide an overview of some current activities at the Census Bureau that relate directly to updating 1990 decennial census estimates of poor schoolchildren, and this overview will then provide the background for the second part of my testimony which covers technical issues and concerns

that arise in launching a program to meet the requirements of H.R. 1645.

Currently, we are working in three areas that bear directly on these questions. The first involves development of specialized poverty tabulations for school districts directly from the 1990 census; the second is the ongoing population estimates program, that we have been doing for a number of years, that updates decennial census population counts for some geographic areas every year, or every two years; and the third is an effort to develop a program for updating 1990 decennial census base estimates of income and poverty, a program that is directly related to the bill under question.

In the first area, we have nearly completed work with the Department of Education to develop estimates of poor school-age children for each of the approximately 15,000 school districts in the country. This effort has yielded two benefits directly related to the requirements of H.R. 1645. First, we have generated poverty estimates for 1989 for school districts, and these estimates are based directly on data from the 1990 decennial census.

Second, working with the Department of Education, we have developed a process that assigns census households to the appropriate 1989/1990 school districts. This process of allocating households to school districts would be necessary under the legislation.

In the second area, we have in place an ongoing program to provide updates of the decennial census population counts for States, counties, and a variety of governmental jurisdictions. Annually, we produce estimates of population for States and counties, and for jurisdictions below the county level we make such estimates of total population on a biennial basis, but we have no system for making population estimates for school districts, a matter that I will return to in a moment.

Evaluations of our estimation methodologies indicate that it is very difficult to make accurate population estimates for small areas—that is, for example, those areas of under 5,000 total population. Our preliminary evaluation of estimates for 1990 based on comparisons with the 1990 decennial census—that is, taking our estimates for 1990 and comparing them with what actually emerged from the census—shows that for areas having populations under 5,000 in 1980, the magnitude of the estimation error for about half of these areas was larger than the actual population change between 1980 and 1990. That is, for very small areas the error in our estimation methods was actually larger than the change over the entire decade, which calls into question the accuracy and the precision of such data for very small areas.

In the third area that I mentioned, we have attempted to establish an income and poverty estimates program to parallel the population estimates program. This project would be undertaken by the Census Bureau as reimbursable work for specific Federal users of these data. In this effort, we have contacted Federal departments and agencies that use income or poverty estimates for purposes of allocating funds, and based on our meetings with these representatives we developed a plan for providing income and poverty estimates that would meet most of their requirements. The plan would be implemented in two phases: first, a phase to develop and test

methodologies; and, second, a phase for producing ongoing estimates on a biennial basis.

Let me now turn directly to the proposed legislation and the issues that can affect our ability to meet the requirements set out in it. First let me emphasize two basic elements needed to fulfill the requirements of estimating the number of poor children by school district. First, we must have estimates of the number of children and the number of families with children by school districts, estimates that we do not now currently produce; and, second, we must have estimates of the income distribution of such families. These are the two basic elements needed to produce the estimates of the number of schoolchildren in poverty in a geographic area.

Population estimates for school districts are an indispensable component in the process of updating decennial census counts of poor children. To meet this need, we would have to expand significantly our current population estimates program. We would need to develop a method for estimating the number of schoolchildren for each school district. We have never made postcensal estimates of population by age for areas below the State level, so that you can see this is a substantial increase in the amount of geographic detail that would be necessary from the State level all the way down to the school district level.

To even attempt population estimates of children by age at the school district level, we would require expanded access to tax return and Social Security information. This would likely include receiving the Social Security numbers for all dependents on tax returns, that information being needed to then obtain information on the age of these children.

The poverty status of children is based on the total income of the family in which they reside. Therefore, to meet the legislative requirements, we must develop estimates of the income distribution for families having school-age children for each school district. The most promising method for estimating family income is based heavily on data from individual income tax returns from the Internal Revenue Service. But using tax returns presents some problems that must be overcome.

For example, not all families file tax returns, especially those with small amounts of income or income from nontaxable sources, and of course those tend to be the families that are of interest when one is examining the poverty population. We must therefore develop methods that bring together the tax information with other data from other sources that better reflect the situation in the very lowest portion of the family income distribution. These sources include local area data on numbers of families receiving nontaxable cash and noncash government benefits—that is, information from administrative record sources for these programs.

I do want to emphasize that we have never attempted to update family income or poverty estimates for detailed levels of geography. There are many unknowns involved, as Chairman Sawyer mentioned a moment ago, in such an undertaking, and we should not assume that all of the problems can be solved in a thoroughly satisfactory manner.

I want to close by mentioning four particular issues that are pertinent to our ability to meet the requirements of the proposed bill.

The first has to do with school district size. As I have mentioned, our past experience with population and per capita income estimation indicates that the accuracy and therefore the usefulness of the estimates declines as the size of the district or the geographic area that we are trying to estimate declines.

Unfortunately for this purpose, most of the 15,000 school districts across the country are very small. About 22 percent of all school districts, I understand, contain fewer than 250 students, and about 74 percent enroll less than 2,500. I doubt very much that we will be able to produce reliable income and poverty estimates for districts this small.

The second is access to tax return data. We have used this in the past for our per capita income estimates, and we will need it; in fact, we will need increased access to such data under the provisions of the bill. But it is my understanding that Internal Revenue Service regulations permit us to receive these files only to support projects which we are directly authorized to conduct and for which we receive direct appropriations under our own title.

The third issue involves our geocoding capabilities—that is, the capabilities of placing households and families and people on the right spot of the Earth, that is, in the right school districts, so that we can associate the right income measure with the right numbers of people. At present, our system for geocoding tax returns to geographic areas does not include coding to the school district level. We hope to upgrade our coding system to provide this capability in a general overhaul of our current methodology that we hope to proceed with, independent of this bill, during this decade. But we don't know now if all of those efforts will be successful, and in any case this new system will not be in place until 1996.

The final issue is the timing of the release of estimates. The bill calls for the release of the first updated poverty estimates in 1995. As we require 15 to 18 months of research and development and additional time after that to implement estimation methods based on the findings of these investigations, I think it would be very difficult for us to meet such a deadline. If we began this fall, I think we could meet a 1996 deadline and that we might be able to produce 1993 data on that mark rather than the 1991 data that is proposed, and then we would catch up over time.

In closing, we anticipate that the small size of most school districts will prove to be an inherently limiting factor in our ability to provide reliable updates of the 1990 census estimates of poor school-age children. Moreover, the methodological and data access issues that must be resolved, even for large geographic areas, are formidable. However, we are optimistic about our ability to make poverty estimates, as specified in the legislation, for counties and areas having populations of 50,000 or more. If the law specifies smaller areas, obviously we will do our best to produce estimates for smaller areas. In summary, I am emphasizing that for areas of 50,000 or more we are confident now, although we do not know exactly the procedures that would be used, that we would be able to do that. For areas less than 50,000, we are less confident. Obviously, we would make our best efforts if called upon to do so.

That concludes my testimony. I will be happy to try to answer any questions.

Thank you.

[The prepared statement of Mr. Butz follows:]

STATEMENT OF WILLIAM P. BUTZ, ASSOCIATE DIRECTOR FOR DEMOGRAPHIC PROGRAMS, BUREAU OF THE CENSUS, U.S. DEPARTMENT OF COMMERCE

INTRODUCTION

Chairman Sawyer and Chairman Kildee, it is a pleasure to be here today to testify before this joint subcommittee hearing on the Census Bureau's population and poverty estimates programs and on the proposed bill H.R. 1645, the "Poverty Data Improvement Act of 1993." In the first part of my presentation I will provide an overview of the current availability of poverty data from the 1990 decennial census and our existing efforts to provide updates of both population and poverty counts for States, counties, and local governmental jurisdictions. This overview will provide the context for the second part of my presentation in which I will focus on the technical issues and concerns that arise in launching a program to meet the requirements of H.R. 1645. I will lay out the steps required in updating decennial census poverty estimates during the 1990s as outlined in H.R. 1645 and the issues that can affect the accuracy and timing of such estimates. While the bill specifies poverty estimates for virtually all governmental jurisdictions, most of my comments relate to the production of estimates for school districts as these areas are the most problematical from a methodological perspective.

As you will see from my testimony, this legislation calls for research that is not at all easy. There are tremendous hurdles to cross to get a system in place for school districts. Unfortunately, crossing these hurdles is not inexpensive or quick, and we are not at all confident that the estimates we would produce for the smallest of these areas would be any better than the estimates currently being used from the 1990 census.

CURRENT STATUS AND PROGRAMS

Currently, we are working in three areas that bear directly on development of a program to update 1990 decennial census estimates of the number of poor schoolchildren. The first involves development of specialized poverty tabulations for school districts directly from the 1990 decennial census. The second is the ongoing population estimates program that updates decennial census population counts annually for States and counties, and biennially for smaller governmental units. The third is a Census Bureau effort to develop an income and poverty estimates program for updating decennial census-based estimates for similar levels of geography.

Specialized Poverty Tabulations

We have nearly completed work with the Department of Education to develop estimates of poor school-age children for each of the approximately 15,000 school districts in the country. This effort has yielded two benefits related to the requirements of H.R. 1645.

First, we have generated "baseline" poverty estimates for 1989 for school districts directly from the 1990 decennial census files. I use the term baseline because any efforts to update poverty estimates during the 1990s will require these data as input.

Second, we have developed a process that assigns census households to the appropriate 1989–1990 school district. We worked closely with the Department of Education in order to enhance our capabilities in this area. For its part, the Department of Education provided detailed maps defining the boundaries of school districts in relation to standard "census" geographic areas [blocks]. The cost of the whole effort, borne by the Department of Education, was in excess of $6 million.

Population Estimates Program

We have in place an ongoing program to provide updates of the decennial census population counts for States, counties, and a variety of governmental jurisdictions. On an annual basis, we produce estimates of population for States and counties. The estimates for States are made for persons by sex within single-year age categories. The estimates for counties are restricted to toal population with no detail on demographic groups. For jurisdictions below the county level [cities, incorporated places, townships, and functioning minor civil divisions], we make estimates of total population on a biennial basis. At the present time we have no system for making popula-

tion estimates for school districts and have only begun rsearch on methods for providing population estimates by age for areas below the State level.

Our population estimation methods rely heavily on administrative records provided by the Internal Revenue Service in the form of tax returns and on data from the Social Security Administration and the Health Care Financing Administration. These data are geocoded using our geographical classification system and merged with data from previous tax years, the previous decennial census, and records on births and deaths to arrive at updates of the decennial census figures.

We conduct periodic evaluations of our estimation methodologies and have found that our estimates for small areas—those places of under 5,000 population—are subject to large errors. For these areas, we see average error levels over a 10-year time span of between 15 and 20 percent. Our preliminary evaluation of estimates for 1990 based on comparisons with the 1990 decennial census shows that, for areas having populations under 5,000 in 1980, the magnitude of the error in our estimate for about half of these areas was larger than the actual population change between 1980 and 1990. Thus, this rate of accuracy is only slightly better than simply using the previous decennial census count for the entire decade.

We have set forth an ambitious plan to improve our postcensal population estimates program during the 1990s. This includes the addition of county population estimates by broad age groups and improved methods for assigning geographic codes [geocoding] to the tax return information that is so vital to the estimation process

Census Income and Poverty Estimates Program

During the past two years, we have attempted to establish an income and poverty estimates program for updating decennial census estimates during the 1990s. This project would be undertaken by the Census Bureau as reimbursable work for specific Federal users of these data. In this effort, we have contacted Federal departments and agencies that use income or poverty estimates for purposes of allocating funds or for other program requirements. This consortium of users expressed a wide range of needs. Based on our meetings with representatives of these organizations, we developed a plan for providing income and poverty estimates that could meet most of their needs.

We have devised a plan that would be implemented in two phases. Since we have never attempted to develop postcensal estimates of household median income or poverty rates for persons as required by members of our consortium, the first phase in our plan would be to test the feasibility of making such estimates. We anticipate a testing period of 15–18 months during which we will examine alternative estimation techniques based on a small sample of States and large counties. If we are successful in our estimation for these large areas, we will extend our research to smaller geographic areas in order to test the relationship between the population size of an area and the accuracy of our estimates. As we have ample evidence from our population estimates program that the accuracy of postcensal population counts for small areas during the 1980s was poor, we must conduct a thorough investigation to understand how the accuracy of income and poverty estimates is affected by population size.

The second phase of our plan to provide updates of 1990 decennial census income and poverty estimates calls for implementation of the results of our research and development work. Guided by these results we would construct a system for producing estimates on a regular, biennial basis. While the ultimate goal of our proposal is to produce estimates for all States, counties, cities, and other incorporated places, we may well not be able to achieve this goal for small areas.

As of this date there are two issues that we must resolve before we can begin work on this reimbursable project. First, we are discussing details regarding the financial contributions that each consortium member will make to funding the project. Second, we are negotiating with the Internal Revenue Service to obtain access to income data contained on their Individual Master File of individual income tax returns. The tax return information is the most important source of data available for estimating the distribution of family income for small geographic areas. I will discuss this critical issue in more detail later.

MEEETING REQUIREMENTS OF THE LEGISLATION

Given this background, I will now turn more directly to the proposed legislation and the issues that can affect our ability to meet the requirements set out in H.R. 1645.

There are three main elements specified in this legislation. First, it calls for the production and publication of estimates for the number of children aged 5 to 17

years old living in families with income below the poverty level. Second, it requires that these estimates be made for all States, counties, local govenmental jurisdictions, and school districts. Third, it specifies that these estimates be made at intervals of not less than two years beginning in 1995.

I will begin by describing the basic elements needed to fulfill the requirements of estimating the number of poor children ages 5 to 17 in school districts. These are [1] estimates of the number of children and the number of families with children by the age of the children, and [2] estimates of the income distribution of such families. I will complete my presentation by pointing to various important issues that must be resolved in developing a system to provide these estimates.

Estimates of Families With Children by Age

Population estimates for school districts are an indispensable component in the process of updating decennial census counts of poor children. To meet this need, we would be required to expand significantly our current population estimates program. We would need to develop a method for estimating the number of children ages 5 to 17 in grades served by each school district. As mentioned earlier, our current plans call only for development of estimates of population by broad age groups at the county level. We would also need to initiate a program to develop estimates of families with children by age of the children because the official measure of poverty is based on the level of family income. We have no experience in making postcensal estimates of the number of families for small areas.

To attempt estimates of children by age at the school district level, we would require expanded access to, and use of, administrative records. This would likely include receiving the social security numbers for all dependents on tax returns. A possible alternative approach might make use of the data on school enrollments collected by the National Center for Education Statistics. Such an approach would have to take into account the number of children enrolled in private schools [about 10 percent nationally], a figure that varies widely among school districts. At present we have no evidence concerning the feasibility of such an approach.

Estimates of Family Income

The poverty status of persons is based on the total income of the family in which they reside. Therefore, we must develop estimates of the income distribution for families having school-age children, for each school district. The appropriate poverty thresholds by size of family can then be used to determine the number of families and number of children in poverty.

While there are other methods that might be used to derive the family income distributions, we think that the most promising are based heavily on Federal individual income tax returns. The advantages of tax returns include [1] ability to geocode tax return addresses to census geographic areas, [2] coverage of a very large proportion of the population, [3] coverage of a very large proportion of the income received by families, and [4] availability as a data source on an annual basis.

Using tax returns also presents some problems that must be overcome. Most notably, not all persons file tax returns, especially those with small amounts of income or income from nontaxable sources. In addition, income defined on tax returns does not include all sources used in the official determination of poverty status and tax filing units are not wholly consistent with our definition of families.

We must, therefore, develop methods that bring together the tax information with other data that reflect the situation in the lowest portion of the income distribution. These sources include local area data on numbers of families receiving nontaxable cash and noncash government benefits. Cash benefits include Aid to Families With Dependent Children and Supplemental Security Income. Noncash benefits include food stamps and free and reduced-price school lunches. While the value of noncash benefits is not counted as income in the official measurement of poverty, participation in these programs is a good indicator of the number of poor. Most of these data are not available for school districts. Hence, our challenge would be to find methods to use the tax and benefit data in ways that lead to estimates of the entire income distribution for all areas needed.

I must reemphasize at this point that we have never attempted to update poverty and income estimates for detailed levels of geography. There are many unknowns involved in such an undertaking and we should not assume that all of the problems can be solved in a satisfactory manner.

Additional Issues

There are additional issues that are pertinent to estimating the number of families and children, and family income distributions. I will discuss them now.

Three of these issues concern the physical characteristics of school districts. These are [1] the small size of school districts, [2] the quality of the information used to code census households to school districts, and [3] the overlapping nature of some school district boundaries.

1. School district size. Our past experience with population and per capita income estimation for small geographic areas shows that the accuracy of our estimates is directly and inversely correlated with the size of the geographic area for which we are making estimates. This relationship also applies to the poverty estimates derived directly from the 1990 decennial census and is revealed in the confidence limits bounding each estimate.

Here are two typical examples. According to the 1990 Census, there were 4,053 children in the grades and areas served by the Highland Park City Schools of Wayne County, Michigan on April 1, 1990. Using data from the long form, collected on a sample basis, the census estimated that 2,003 of these children were living in families with income below the poverty level for 1989. The width of the 95 percent confidence interval for this estimate is 500 children, ranging from 1,705 to 2,301. The second example from the Coventry Public School system of Summit County, Ohio [1,800 school-age children] shows an estimate of 175 poor children in 1989 and a 95 percent confidence interval 184 children wide—between 83 and 267.

These wide confidence limits on our estimates of poor children reflect large sampling variability because these estimates are based on the Census sample and the size of the sample is proportional to the size of the area. Given the degree of uncertainty associated with the census estimates, we can anticipate significant problems with estimates from update methodologies as well.

Unfortunately for this purpose, most of the 15,000 school districts across the country are very small. About 22 percent of all school districts contain fewer than 250 students and about 74 percent enroll less than 2,500. I doubt very much that we will be able to produce reliable income and poverty estimates for those districts.

2. Quality of the school district boundary data. The quality of school district boundary data is very important to the overall quality of population and poverty estimates. We have experienced some problems in assigning census households to grades in school districts. These problems occur mainly in situations where the school district boundary "split" a census block. Better methods need to be established for allocating the children to the appropriate school district.

We would also need to develop an ongoing system to update the geocoding system for changes in the boundaries of school districts. If such work is not undertaken, our estimates could reflect only the situation as of the 1989–1990 school year. Our information indicates that school district boundaries change for between 100 and 200 areas each year.

3. Overlapping school district boundaries. School district boundaries sometimes overlap and these overlaps may complicate the estimation process. This overlap affects about 25 percent of the total number of school districts. In these cases of overlap, children in the same family may attend school in different school districts depending on their grade level. We must therefore devise methods that account for these overlaps when computing estimates of poor children.

Finally, I have four other concerns that I would like to mention: [1] access to the tax return data needed in the estimation process, [2] enhancements needed in our geocoding system, [3] need for a formal estimates review process, and [4] completion of the first round of estimates in 1995.

1. Access to tax return data. As I mentioned earlier when discussing our small areas income estimates proposal, we will need access to an extract of information from the Internal Revenue Service's Individual Master File of all Federal individual income tax returns. This will permit us to geocode these returns to school districts and to test various estimation strategies. It is our understanding that Internal Revenue Service regulations permit us to receive this file only to support projects which we are directly authorized to conduct and for which we receive direct appropriations.

Although H.R. 1645 would amend section 182 of Title XIII, U.S. Code, to authorize the Bureau of the Census to produce and publish poverty estimates for school districts, the Congress may decide to provide funding for this work to another agency, such as the Department of Education. As we are not permitted to use the tax return data to perform work under remibursable agreements with other agencies, we may be denied access to the data for the purposes specified in the bill. Our ability to make the estimates requested in an efficient and timely manner could be diminished under such an arrangement.

2. Geocoding capabilities. At present our system for geocoding tax returns to geographic areas does not include coding to the school district level. We hope to up-

grade our coding system to provide this capability in a general overhaul of our current methodology. The new system will provide this capability for areas having city-style street addresses but will not be able to assign school district codes for areas where the addresses are not city-style. We do not anticipate that this new system will be in place before 1996. Until that time, we will be limited to current procedures that can assign codes covering States, counties, and other incorporated places.

3. Establishment of a review and challenge process. Because the poverty estimates will be used to allocate funds, we will need to establish a system that permits school districts and other stakeholders to review our procedures and challenge the validity of our estimates. Such a review and challenge process was an integral part of our efforts to provide population and per capita income estimates for the General Revenue Sharing program in the 1970s and early 1980s. For that program we received between 500 and 1,000 formal challenges of our estimates annually. Given the anticipated difficulty in making poverty estimates for school districts and the levels of funding involved, we expect that a formal review process could generate a much larger volume of challenges.

4. Timing of the release of estimates. The bill calls for the release of the first updated poverty estimates in 1995. As we would require 15–18 months of research and development and additional time to implement estimation methods based on the findings of our investigations, it would be very difficult for us to meet such a deadline unless work on the program could begin within the next several months.

CLOSING

We anticipate that the small size of most school districts will prove to be an inherently limiting factor in our ability to provide reliable updates of the 1990 decennial census estimates. Moreover, the methodological and data access issues that must be solved in order to meet the requirements of this legislation, even for large geographic areas, are formidable. However, we are optimistic about our ability to make poverty estimates as specified in the legislation for counties and areas having populations of 50,000 or more.

That concludes my testimony. I will be happy to answer any questions.

Chairman KILDEE. Thank you very much.

Mr. Elliott.

Mr. ELLIOTT. Thank you very much, Chairman Sawyer and Chairman Kildee. I have a prepared statement that has been submitted, if I could have that in the record, and I will do a summary.

Chairman KILDEE. Yes, it will be included in its entirety.

Mr. ELLIOTT. Thank you very much.

To complement the things that Mr. Butz has said, I would like to address my comments this morning to five questions in particular: Why were the school district estimates developed to begin with? How are the estimates developed? How are the estimates used in the Chapter 1 program, Chapter 1 being the largest Federal assistance to elementary and secondary education? How does the reliability of the estimates affect the distribution of Chapter 1 funds? And then, finally, how might the estimates be improved that are used for allocating Chapter 1 funds?

First, why were the school district estimates developed to begin with? As a part of the 1988 Hawkins-Stafford Elementary and Secondary School Improvement Amendments, Congress mandated that the National Center for Education Statistics submit a report after each decennial census on the social and economic status of children who reside in local education agencies, and Mr. Butz has just described a part of that process. The estimates of related children 5 to 17 years in poverty families by school district is needed for the Chapter 1 program, and it is one of several products from this overall project.

There are, I think, at least four reasons that these kinds of estimates are needed. One is that, in fact, the data are used as a basis for allocating Federal funds, including funds in Chapter 1. Second, there is a within-State use of the data, so that the Federal Government, if it allocates down to counties, still provides additional information through the school district data so that States can use it to allocate among the school districts within a county within their own State. Third, things do change from one decennial census to the next decennial census, and they need to be brought up to date.

And finally, Mr. Kildee—I think my recollection is correct on this—there were a number of States, led by Michigan actually, in 1987 that were very much concerned when it looked as though the National Center for Education Statistics and the Bureau of the Census would not be able to provide the decennial census data by school district as they had for 1970 and 1980. Michigan and a group of other States were very intent on using the data for within-State planning purposes. It was very important to them, and they made sure the Congress understood that in 1987 when our legislation was under consideration.

Second, how were these estimates developed? The Census Bureau, in conducting the 1990 census, used two questionnaires to enumerate households. They used a short form containing basic demographic and housing information for most households in the country. In addition, they used a longer form with many more questions, including questions on income, for a sample of households. Estimates of children ages 5 to 17 years and families with income below poverty are developed from households that received the long form or the sample questionnaire.

Third, how are the estimates used in the Chapter 1 program? The Department of Education allocations for Chapter 1 are computed on the basis of county level data. States are responsible for making suballocations down to the school district level, and States use a variety of data to make the district allocations. Some States use the census estimates of children in poverty to allocate among districts. Others use items such as the number of children in the school lunch program or the AFDC children to allocate funds among the individual school districts within a county. These latter items have an advantage in that they can be kept current. States also have the responsibility for making corrections and adjustments to the data to reflect boundary changes in school districts, such as consolidations.

Fourth, how does the reliability of the estimates affect the distribution of Chapter 1 funds? When the Census Bureau, the National Center for Education Statistics, or other statistical agencies release estimates to the public, they typically provide a discussion of sources of error that might affect the data as well as their suitability for various purposes. These discussions usually differentiate two kinds of error: sampling error and nonsampling error or bias. Nonsampling error that may affect these estimates can occur as a result of several different kinds of problems. For example, the district boundaries may have been drawn incorrectly, or a digitizing error may have occurred when the Census Bureau computerized the boundaries of the paper maps. Another example: districts that

have a higher proportion of their population in split blocks within the census count will have lower reliability.

But for most districts sampling error is probably the more significant source of error affecting the estimates. As noted earlier, these estimates are derived from a sample that was part of the 1990 decennial census. Nationwide, one out of every six households was asked the income questions that are used to determine poverty status. For the 1990 census, the Census Bureau used higher sampling rates for small governmental units to increase the reliability of the estimates for those areas. The minimum sample size was one in eight for tracts in large cities, and the maximum was one in two for small incorporated towns. Because school districts were not considered in the decisions about sampling, the effective sampling rate for school districts ranges from one extreme to the other. Small school districts, however, tend to be associated with small towns or rural areas that were sampled at a higher rate.

We believe that most of the challenges of the 1990 census estimates made by States and school districts reflect estimates that are strongly affected by sampling error. Typically, they would be cases where the estimate is significantly lower than the true value. Such cases have a low probability of occurring. The fact that they do occur, however, is one of the shortcomings of using estimates from probability samples for the purpose of distributing money to small jurisdictions.

Poor children in small districts are much more likely to receive too little or too much Chapter 1 money than are poor children in large districts because the sampling variability in the estimates is greater in small districts. Although this misallocation of funds can have a significant effect on the program of an individual district, it will likely have a negligible effect on the overall targeting of Chapter 1 funds. This is because the vast majority of cases that have a large percentage error will occur in the small districts and therefore the error will affect relatively few children.

Fifth, how might the estimates be improved? Probably the best procedure is one that was used by the Census Bureau to improve the estimates of per capita income for the general revenue sharing program. The per capita income estimates from the decennial census had the same problem that we face here—that is, large sampling errors for small jurisdictions. The procedure, called James-Stein estimates, is a method of incorporating additional information into the estimation process, especially information with little sampling errors, so that the resulting new estimates are closer to the true value than the original statistics. Thus, application of the James-Stein procedure could result in considerable improvement in the estimate.

Information such as the number of children receiving free or reduced-price lunch or the number of AFDC children could be a good predictor of the number of children in poverty even if they do not measure precisely the same thing. NCES is currently discussing the possible application of this procedure to develop more accurate estimates with several States.

A second way of improving the estimates would be to update them for the changes that occur over time. Although it would be most useful if the Census Bureau could develop periodic estimates

for all counties and districts, as proposed in Congressman Sawyer's bill, the Bureau's testimony indicates that updates of the district data would be of questionable quality. The use of updated estimates for counties would go a long way toward keeping Chapter 1 allocations current with changing conditions. County updates of children in poverty would permit updating the allocations for States, and county estimates could also be useful as controls in State efforts to update children in poverty estimates.

States, perhaps working with the National Center for Education Statistics, may be in a better position than the Census Bureau to make such updates since they are able to make use of local and State sources of data as well as national sources. If the county allocations are updated, I would recommend encouraging States to explore the feasibility of developing such updates for their districts. The national Center for Education Statistics would be happy to provide technical assistance to the States as needed. We would be prepared to answer your questions.

Thank you, Mr. Chairman.

[The prepared statement of Mr. Elliott follows:]

Statement of Emerson J. Elliott

Introduction

Chairman Sawyer and Chairman Kildee, thank you for the opportunity to testify at this joint subcommittee hearing. In response to the stated focus of the hearing and discussions with Committee staff, I will limit my remarks to five aspects of the development and use of the 1990 Census estimates of children in poverty for the Chapter 1 allocations. The five aspects are:

Why were the school district estimates developed?
How were these estimates developed?
How are the estimates used in the Chapter 1 program?
How does the reliability of the estimates affect the distribution of Chapter 1 funds?
How might the estimates be improved?

Why were the school district estimates developed?

As part of the 1988 Hawkins-Stafford Elementary and Secondary School Improvement Amendments,P.L.100-297, Congress mandated NCES to "submit a report to the appropriate committees of the Congress concerning the social and economic status of children (based on data from the most recent decennial census) who reside in areas served by different local educational agencies". The estimates of related children 5-17 years in poverty families by school district, needed for the Chapter 1 program, is one of several products of this overall project. For the purposes of this testimony, my discussion is limited to this use of the data.

How were these estimates developed?

The Census Bureau, in conducting the 1990 census, used two questionnaires to enumerate households. They used a short form, containing basic demographic and housing information, for most households in the country . In addition, they used a long form or sample questionnaire, which comprised the short form items as well as many others--including questions on income-- for a sample

of households. Estimates of children ages 5-17 years, in families with income below poverty, are developed from households that received the long form (sample) questionnaire.

The process used to develop the estimates for school districts involved numerous steps:

1. State officials drew school district boundaries on maps supplied by the Census Bureau.
2. The Census Bureau "digitized" the boundaries. This process creates computer-readable school district boundaries.
3. The Census Bureau assigned all households with children ages 5-17 years, which completed a 1990 census sample questionnaire, to a school district as follows:
   a. Households were classified by block in the census.
   b. Blocks wholly contained within a district were assigned to that district.
   c. Blocks with a school district boundary running through them were split into two (or more) parts. Households in the split blocks were allocated to the respective school districts based on specifications provided by the State. Typically, this was done proportionally to land area.
4. NCES determined the grade range for each school district based on information provided by the State as part of NCES's 1989-1990 Common Core of Data collection.
5. The Census Bureau determined the grade for each child, 5-17 years, based on the respondents responses to the census question on highest grade completed.
6. Children were classified as "relevant" for a particular school district if they lived in a housing unit within the district's boundaries and their current grade was within the district's grade range.

7. Relevant children were classified as "in poverty" if their family's pre-tax cash income, as reported on the sample questionnaire, was below the official poverty threshold for that size and type (e.g. one parent, two parents) of family.
8. The Census Bureau developed a "weight" for each person in the sample such that the sum of the weights was equal to the total census count. The Census estimate of children 5-17 years of age in poverty families is therefore equal to the sum of the weights of such children in poverty families in the sample.

How are the estimates used in the Chapter 1 program?

School district level data are not used by the Department of Education to calculate Chapter 1 allocations for States. Departmental allocations are computed on the basis of county level data. States are responsible for making sub-allocations down to the school district level, and States use a variety of data to make the district allocations.

Some states use the Census estimates of children in poverty to allocate to districts. Others use items, such as the number of children in the school lunch program or AFDC children, to allocate the funds to individual school districts. These latter items have an advantage in that they can be kept current. States also have the responsibility of making corrections and adjustments to the data to reflect boundary changes, consolidations, etc.

How does the reliability of the estimates affect the distribution of Chapter 1 funds?

When the Census Bureau, NCES, or other statistical agencies release estimates to the public, they typically provide a discussion of sources of error that might affect the data as well

as their suitability for various purposes. These discussions usually differentiate two kinds of error--sampling error and non-sampling error or bias. Non-sampling error that may affect these estimates can occur as a result of several different kinds of problems:

First, the district boundaries may have been drawn incorrectly or a digitizing error may have occurred when the Census Bureau computerized the boundaries from the paper maps.

Second, districts that have a higher proportion of their population in split blocks will have less reliability. This source of error was much diminished for 1990 compared to 1980 because of the much smaller block size in 1990. There were 7 million blocks in 1990 vs. 258,000 in 1980.

Third, there was also less error in assigning children to districts in 1990 in situations where there are overlapping districts. For 1990, the assignment of children to elementary vs. secondary districts was made using the child's enrollment grade rather than the child's age, as was done in 1980. The use of age is known to introduce a bias favoring secondary districts. For example, in 1980, 14-year-olds were typically allocated to secondary districts even though at the time of the census about one half of 14-year-olds were still in the 8th grade and were the responsibility of an elementary district.

Finally, there are various errors associated with the taking of the census. For example, the census may fail to count some people or respondents may fail to report their income correctly. On balance, respondents to Census surveys report less income than administrative records show they receive. This underreporting of income is especially common among households at the high and low ends of the income distribution. Combined with the practice of excluding from poverty calculations the considerable amounts of noncash income households receive, this underreporting of pre-tax

cash income would tend to increase the number of poor children aged 5-17 in the sample. We have no reason to think that underreporting of cash incomes by low-income households would be different in small districts than in large. So, the effect of income underreporting probably would be independent of the sampling errors discussed below.

For most districts, sampling error is probably the most significant source of error affecting the estimates. As noted earlier, these estimates are derived from a sample that was part of the 1990 Decennial Census. Nationwide, one out of every six households was asked the income questions that are used to determine poverty status. For the 1990 Census, the Census Bureau used higher sampling rates for small governmental units to increase the reliability of the estimates for those areas. The minimum sample size was 1 in 8 for tracts in large cities and the maximum was 1 in 2 for small incorporated towns. Because school districts were not considered in decisions about sampling, the effective sampling rate for school districts ranges from one extreme to the other. Small school districts, however, tend to be associated with small towns or rural areas that were sampled at a higher rate.

As a result of sampling, the estimates are subject to random variability inherent in all sampling processes. For that reason, the estimate may be too high or too low compared to the "true value" that would have been obtained if all households had been asked the income questions. In order to provide a basic understanding of this random component of error, I will describe the situation where this type of error is most likely to be serious, that is, in very small school districts. In this example, let us assume that there are 100 school districts, each having an enrollment of 300 children, of whom 20 percent, or 60 children, are in poverty. If all households in a census were asked the income questions, the Census would estimate 60 children in poverty for each district. If, however, only one in three

households were asked the income questions, the estimates would vary from district to district. According to statistical theory, about one half of these districts will have estimates within 15 percent of the true value (i.e. between 51 and 69), and about two-thirds will have estimates within 22 percent of the true value (i.e. between 47 and 73). On the other hand, the estimates for the remaining one third of the districts will be more than 22 percent from their true value and about 5 percent will be off the mark by over 50 percent (i.e. less than 30 or greater than 90).

As districts increase in size, this distribution becomes narrower and the estimates ,on average, become more reliable.
We believe that most of the challenges of the 1990 Census estimates made by States and school districts reflect estimates which are strongly affected by sampling error. Typically, there would be cases where the estimate is significantly lower than the true value. As we have shown above, such cases have a low probability of occurring. The fact that they do occur, however, is one of the shortcomings of using estimates from probability samples for the purpose of distributing money to small jurisdictions. Poor children in small districts are much more likely to receive too little or too much Chapter 1 money than are poor children in large districts, because the sampling variability in the estimates is greater in small districts. Although this misallocation of funds can have a significant effect on the program of an individual district, it will likely have a negligible effect on the overall targeting of the Chapter 1 funds. This is because the vast majority of cases that have a large percentage error will occur in small districts and, therefore, the error will affect relatively few children.

Table 1 shows the size distribution of school districts in the U.S. Overall, about 50 percent of all districts have less than 1000 students, and they account for a little over 7 percent of all students attending public school. The smallest 25 percent of all districts have less than 300 students. All of these districts

combined, however, enroll only 1.3 percent of all school children. Thus, it would appear that the use of the census estimates provides a "correct" distribution of funds overall, but that a significant number of small exceptions occur. In the next section, I will discuss the question of what might be done about these exceptions.

How might the estimates be improved?

As indicated above, sampling error is the major source of error for most of the school district estimates. In a way, this is fortunate because there are techniques that can be used to reduce this error. Probably the best procedure is one that was used by the Census Bureau to improve the estimates of per capita income (PCI) for the General Revenue Sharing Program. The PCI estimates from the decennial census had the same problem that we face here--large sampling errors for small jurisdictions.

The procedure, called James-Stein estimates, is a method of incorporating additional information into the estimation process--especially information with little sampling error--so that the resulting new estimates are closer to the true value than the original statistics. The procedure has its largest effect on estimates with large sampling error. The amount of the improvement depends upon how good the additional information is in predicting the true value for each district. In our case, information, such as the number of children receiving free or reduced price lunch or the number of AFDC children, could be a good predictor of the number of children in poverty even if they do not measure precisely the same thing. Thus, application of the James-Stein procedure could result in considerable improvement in the estimates. NCES is currently discussing the possible application of this procedure to develop more accurate estimates with several States.

A second way of improving the estimates would be to update them for changes that occur over time. The Census Bureau's testimony indicates that periodic updates for all districts, as proposed in H.R. 1645, would be of questionable quality. Although district level updates would be useful, they are not critical, in my opinion.

The use of updated estimates for counties would go a long way toward keeping Chapter 1 allocations current with changing conditions. County updates of children in poverty would permit updating the allocations for States. In addition, updated county controls could be carried down to districts for States that use items like children in the school lunch program when making sub-county allocations to districts.

County estimates could also be useful as controls in State efforts to update children in poverty estimates. States, perhaps working with NCES, may be in a better position than the Census Bureau, to make such updates since they are able to make better use of local and state sources of data. If the county allocations are updated. I would recommend encouraging States to explore the feasibility of developing such updates for their districts. NCES would be happy to provide technical assistance to the States as needed.

I would like to thank the Committees for this opportunity to testify and will be pleased to answer any questions you may have.

Chairman KILDEE. Thank you very much.

Ms. Ruggles.

Ms. RUGGLES. Thank you. I am happy to be here today to discuss the provision of updated income and poverty estimates for geographic areas at the State level and below.

I plan to address three main questions: Do we actually need more frequent poverty estimates? Is it feasible to do more frequent poverty estimates for smaller geographic areas? And what data and other resources would be needed in order to undertake this kind of estimate? Although these questions arise in considering H.R. 1645, the Poverty Data Improvement Act of 1993, I plan to focus on the general issues involved rather than on the specific provisions of that bill. I will, of course, be happy to answer any questions you may have.

Poverty estimates are used for a number of different purposes. For example, they may be used to track the relative well-being of different demographic groups within the population, to assess and compare the impacts of various economic trends over time, and, not least, to allocate funds under specific Federal programs to aid the poor.

The Census Bureau currently produces two major types of poverty estimates that are used for these purposes. First, estimates of the total number of poor in the United States as a whole are produced every year using data from the Current Population Survey, a survey of about 60,000 households. These estimates provide the basis for the poverty rate figures that Census releases every year.

Although the CPS is quite large as sample surveys go, it doesn't include enough households to be representative for each State individually. Reasonably reliable estimates of poverty rates can be made from this survey for some of the larger States. For smaller States, however, the survey does not collect information on enough different households in each State to provide an accurate picture of income and poverty status for the State population as a whole.

Estimates for counties and other jurisdictions below the State level generally cannot be made using CPS data. Instead, poverty estimates for jurisdictions at the State level and below are produced using data from the decennial census.

The decennial census collects information on many more families than does the CPS, and, as you know, every household in the country returns a census form or is supposed to return a census form. Estimates of income and poverty status are computed using the long form returns, which are completed by about one household out of six. The census long form collects the basic information on sources of family income and on household composition that are needed to determine each family's poverty status. In all, about 15 million long forms were completed for the 1990 decennial census. This very large sample of returns allows income and poverty estimates to be calculated for even very small geographic areas, including counties, cities, and towns.

The major advantages of the decennial census are its very large size and its comprehensive coverage of all areas of the country. Its major disadvantage, of course, is the fact that it is performed only once every 10 years. As we know from the CPS and other annual surveys, poverty rates can fluctuate quite a bit from year to year.

Roughly 30 percent of the people who are poor in any given year will not be poor the next year, but new families will be entering poverty instead.

Across the Nation as a whole, differences in the numbers of people leaving and entering poverty lead to changes in the overall poverty rate. When there are more exiters than entrants, the rate goes down, and when there are more entrants than exiters, it goes up. Even at the broad national level, poverty rates are seen to fluctuate by as much as two to three percentage points within a fairly short period of time. Between 1989 and 1991, for example, the overall poverty rate rose from 12.8 percent to 12.42 percent of the U.S. population. This represented an increase of four million in the number of people in poverty. So that is a pretty big change over a two-year period.

These changes in poverty rates are not equally distributed across all geographic areas. One region of the country may be particularly hard hit by a recession, for example, while another is experiencing a relatively high rate of job growth. Other factors, such as the migration of poor people within an area or the influx of poor immigrants into the area, can also affect local poverty rates. Over a period of 10 years, local jurisdictions may experience widely varying changes in their poverty rates.

Because local area poverty statistics are computed from the decennial census, they are available only once every 10 years, and because it takes some time to collect and process data on income and population, these estimates are typically already a few years out of date when they become available. For example, the most recent set of estimates which relates to family incomes received in 1989, just became available for local jurisdictions this year. Without some special effort to update these estimates, they will be in use from now until approximately the year 2003.

These estimates are already somewhat out of date. The past four years have seen some substantial changes in income and poverty rates for the Nation as a whole, and these changes have affected different jurisdictions quite differently. It is not hard to imagine that by the year 2003 the distribution of poor people across different jurisdictions within the United States could be very different from what it was in 1989. Allocation of funding for Chapter 1 and similar programs will not necessarily go to the jurisdictions with the largest poverty problems in the year 2003 if these allocations are based on the geographic distribution of poverty that we saw in 1989. Basing allocation formulas on data that are this outdated seems neither efficient nor fair.

What alternatives could be explored to improve the quality of our income and poverty estimates for local areas? Because there are large differences in levels of need across different counties, cities, and so forth within the United States, it does seem that it is necessary to use some kind of indicator of local area income in deciding how to allocate funds. The use of data based on the 1989 distribution of needs allows for very precise targeting of funds, but unfortunately that precision may not be closely related to actual needs and, in fact, is likely to become less accurate as time goes on. Can these 1989 income and poverty estimates be updated in some meaningful way before the next decennial census?

As Mr. Butz explained, the Census Bureau does update its estimates of the populations of specific jurisdictions, such as States, counties, and cities, between decennial censuses. Using data from tax and Social Security records as well as vital statistics reports, updated estimates of the total population are produced for States and counties every year and for subcounty jurisdictions every other year. These estimates rely on baseline population counts from the decennial census, but more recent data are used to project trends in population since the census within each specific jurisdiction.

These estimates do not contain a high level of detail. State estimates are available by age and gender, and only population totals are available for most other jurisdictions. Nevertheless, these estimates are substantially more accurate for reasonably large jurisdictions—over 50,000 in population, for example—than estimates based on the decennial census alone would be. Because the tax and wage records used in updating are not comprehensive, however, estimates for very small jurisdictions, which may not be well represented in the available records, are substantially less accurate.

Currently, Census does not attempt to track year-by-year changes in income and poverty status across jurisdictions at the State level and below. Recently, however, the Bureau has proposed using a methodology similar to that used to update the population estimates to produce updated income and poverty statistics for specific jurisdictions. Such a procedure would be feasible.

The same tax and Social Security records that are used to update the population estimates also contain information on incomes. This income information could help to identify trends in income and poverty status over time across different local jurisdictions. If wages were stagnating and unemployment was high in a particular area, for example, the impacts of these problems would be clear in the tax and Social Security data, and that information could be used in updating the poverty estimates.

Conversely, areas where wages were high and jobs plentiful could also be identified. Trends in poverty among those not in the labor force—for example, the elderly and the disabled—are harder to track using tax and wage data, but these data could be supplemented with information from other sources, such as the food stamp program.

As with the updated poverty estimates currently performed by Census, only the year-to-year trends would be tracked across jurisdictions using the administrative data. Baseline estimates of income and poverty status would still need to be calculated from the decennial census. These estimates would simply be updated between censuses using other sources of data. When new decennial census data became available, the estimates would need to be rebenchmarked for each jurisdiction in much the same way that the population estimates are currently recalibrated. Nevertheless, as with the population estimates, the updated income and poverty estimates would almost certainly prove to be more accurate for jurisdictions over 50,000 in population than are estimates based on the decennial census alone.

After reviewing the available data, in other words, I believe it would be both useful and feasible to update income and poverty estimates for States, counties, and other jurisdictions over 50,000 in

population on a regular basis between decennial censuses. What, however, should we do about jurisdictions with populations below 50,000?

The decennial census provides very detailed information on geographic areas, and reasonably reliable income estimates can be made using these data for jurisdictions as small as 5,000 in population in the year of the decennial census. Such a jurisdiction would typically be represented by more than 800 long form returns in the decennial census, enough to get a representative picture of the income distribution. Unfortunately, no other source of data available to the Census Bureau contains such a wealth of information on the incomes received within such small jurisdictions, however.

Updating income and poverty estimates requires a relatively large sample from each jurisdiction covered because it is necessary to have representative records across many different income categories if the updates are to be accurate. The Census Bureau's own studies have shown that they cannot reliably estimate even population changes between decennial censuses for places below 5,000 in population because there just isn't enough detailed information available.

Because updating income and poverty estimates is more complex than updating population estimates and requires even more detailed data, it is very unlikely that accurate and reliable estimates can be produced for smaller jurisdictions. School districts, for example, typically have populations of less than 5,000 and are therefore not good candidates for individualized income and poverty updates.

What are the implications of this problem for local area poverty estimates? The fact that we cannot produce specific update factors for smaller jurisdictions doesn't mean we can't use any data at all on these jurisdictions, it just means that the best estimate of the change in income and poverty rates experienced in these jurisdictions is likely to be the change observed for the larger jurisdiction, such as counties, cities, and towns, of which they are a part.

Because we do have baseline data from the decennial census on relatively small jurisdictions, including many school districts, major differences in their original distributions of income can be taken into account when considering funding allocations. In updating those estimates, however, the best we can do between censuses is to project income and poverty status assuming that the direction and magnitude of change is similar to that seen in the county or the city as a whole. Attempting more detailed updates than this might increase precision, but it would do little to increase the accuracy of the estimates.

In conclusion, I would just like to say that the decennial census long forms are our major source of information on incomes and poverty status. These census data are very comprehensive and provide a lot of geographic detail, and they do allow us to make more accurate estimates of income and poverty status across States, counties, and other jurisdictions. However, they do become fairly outdated pretty fast, especially in times of rapid economic change. A systematic series of updates between the decennial censuses could greatly improve the accuracy of these estimates, especially for jurisdictions over 50,000 in population.

75-540 O - 94 - 2

The Census Bureau's proposed methodology for performing such updates appears feasible and appropriate. Because it relies heavily on administrative records that are already being collected for other purposes, it would also be reasonably cost effective, although some increase in funding would be needed to implement this program. Given the importance of the issue and the large sums of money allocated using income and poverty data, however, such a funding increase would be a worthwhile investment. I commend the Bureau and these two subcommittees for your proposals to improve the quality and the timeliness of these data.

Chairman KILDEE. Thank you very much, all of you, for your testimony.

This committee will be reauthorizing Chapter 1 this year, and we find a number of difficulties. Without getting into all of them, some are political, some are statistical, some are fiscal, there are a variety. We may have to try to massage those in the fairest way possible. States in the Northeast, particularly States like Massachusetts and Rhode Island, seem to have had this situation when the census was taken. The good news was, they didn't have as much poverty. The bad news was, that drove relatively fewer dollars to those States for Chapter 1. In Michigan, on the other hand, the bad news was, we had a lot of poverty. The good news was, that drove relatively more dollars to the State of Michigan. Those are some of the things we have to wrestle with, and those are more political matters.

But it is important for us in the next decade, and maybe even this year since we are somewhat removed yet in time from when the census was taken—it is important for us to have at our disposal as up-to-date data as possible so we can make sure that the purpose of Chapter 1 is carried out, that we actually help kids.

What would the additional cost be to develop mid-decade poverty estimates for counties?

Mr. BUTZ. Mr. Kildee, let me try that one. In order to give you a precise estimate, we would really have to take a look at the numbers and get back to you, but the methodology that would be required is the same—the development of methodology is the same whether one were going to do it once or do it every other year.

So I think part of the answer is that for the first two years or so of development it really doesn't matter, and the bill has funding in it for that, and if that bill were passed we would proceed on that basis. In order to produce production numbers—that is, at the middle of the decade—would require somewhat more; how much more I don't know at the moment, but we could work up an estimate—that is, the actual annual amount to produce on a production basis.

Chairman KILDEE. But when you are doing it for a specific purpose, to determine poverty, you could zero in on that part of the data you are gathering.

Mr. BUTZ. Yes, sir.

Chairman KILDEE. So it would not be a question of replicating the entire census over again, you would be looking for some particular——

Mr. BUTZ. No, that is correct. The update would be done not on the basis of taking another census but on the basis of using infor-

mation from the previous census—in this case, the 1990 census—in combination with information from various administrative record systems—for example, the tax data, AFDC case records and the like—in order to update the information, but it wouldn't require additional primary data collection of a census sort.

Chairman KILDEE. Very good.

Mr. Sawyer.

Mr. SAWYER. Thank you, Mr. Chairman.

Mr. Butz, you talked at one point in your testimony about the lack of confidence in data for small areas, under roughly 5,000. You spoke at another point in terms of your confidence in the ability to produce estimates over 50,000. What do we do in between?

Mr. BUTZ. Well, that 50,000 cut-off I know sounds very arbitrary, but it is an attempt to come up with an approximate number. That is the level above which, based on our research, we are very confident that we can produce good data—that is, data that are definitely better as the decade proceeds than using information from the last census. That 5,000 cut-off was just simply an illustration. At that cut-off, about half of the very small jurisdictions for which we do population estimates had more error than the 1980 to 1990 change.

As you go up from 5,000 to 50,000, in general, overall, the quality of the estimates improves, but it isn't until you get to about 50,000 that we are confident that we will be able to develop measures that will produce better information. So the 5,000 is terrible; the 50,000 we are confident will be good; and below 50,000 we are really not sure that it will be better for enough areas, whether it be counties or school districts, to make the effort worthwhile.

Mr. SAWYER. Which is it more important to get the numbers right for, the largest number of school districts in the country, which would be below 5,000, or the largest number of schoolchildren in the United States, who are in districts above 5,000?

Mr. BUTZ. Well, I can't answer the question of which would be most important, but certainly if we do the estimates for areas above 50,000, most of the children are in larger school districts, so that the distribution of children by school districts is very different from just the distribution of school districts. I don't know, in fact, what proportion of children are in very large school districts, but certainly a much larger proportion are than the proportion of such school districts of the total.

Mr. SAWYER. Mr. Elliott supplied a table that is awfully interesting in those terms. I would commend it to all of the members of the committee. It really places in perspective that relationship between districts and children.

Let me ask you this, because this is really where I am leading, and perhaps more than one of you could comment on this, about the potential to merge in similar areas, in areas of over 50,000, like-structured districts with demographic and economic similarities that you could measure on a fairly large scale, and to develop from those merged territories greater accuracy in terms of intercensal projections and estimates than you could simply by using the originally developed data alone.

Was that a clear question?

Mr. BUTZ. Yes, sir, and certainly that would be a help. The larger the district, whether it comes from combining small ones or whatever, in general, the greater the accuracy of the estimates.

Ms. RUGGLES. A variance on that would be simply to use data for the town or the city in which the school districts were located as the basis for updating the estimates. You would still use the basic school district base to get a baseline for what the poverty level looked like in different places, but in updating that you would just assume that there was about the same level of economic growth throughout the town or throughout whatever the jurisdiction was.

Mr. SAWYER. I can appreciate that. I am always concerned, though, about using averages over a fairly large area where there are, in fact, very substantial differences in terms of local jurisdictions within the larger county. I know of a number of members of this committee who experience phenomena of that kind.

Let me ask just a couple of very quick questions. You mentioned the problem of nontax filers. What other kinds of administrative records would serve as useful proxies in trying to improve those estimates for nonfilers?

Mr. BUTZ. I think the kinds of data that we would look at would be AFDC rolls, food stamp rolls, case rolls of children who participate in free or reduced-price school lunches, that kind of information, to try to get at the segment of the population that is most likely to be in poverty and whose families are least likely to file IRS returns. Which sources we would actually use and how we would combine or model them would be a question for research, but it would be those kinds of data that would be used, and they would indeed be very, very important because we can't just rely on IRS tax return data to do this job.

Mr. SAWYER. Mr. Chairman, I have really only one more. I don't want to monopolize the time.

Mr. Elliott, I understand that the tabulations for school-age children by school district won't be completed until 1994 or 1995. We have received a number of inquiries about that. I was wondering if you could comment on that.

Mr. ELLIOTT. We are expecting the data to be available by October of this year, Mr. Chairman.

Mr. SAWYER. Is that correct?

Mr. ELLIOTT. That should be available for all districts in all States by that time.

Mr. SAWYER. That is good news.

Mr. ELLIOTT. May I pick up just a bit on the question that you were posing earlier about, should we try to cover more districts or more students?

Mr. SAWYER. I would be pleased to have you comment.

Mr. ELLIOTT. That is a sort of question that I think a statistical agency can't answer, but I would like to make a comment about it, and that is, to some extent the issue is pivotal around what kinds of other data are available that might be used to go to some level below the county and whether or not those data are up to date and satisfactory.

Now we have used extensively in States information on school lunch and AFDC, but that judgment about whether that is a satisfactory substitute really has to be a political judgment.

Mr. SAWYER. Mr. Elliott, I hope you understand that the purpose of the question was to get at the capacity to provide appropriate mergers for districts for which you might find other symptomatic measures or proxies, whether it be any of those that have been mentioned, or race, or ethnicity, or any of the others that would allow merger of like areas into sufficient size to provide the level of confidence that you need to drive dollars based on that data.

Mr. ELLIOTT. Right.

Mr. SAWYER. Thank you.

Thank you, Mr. Chairman.

Chairman KILDEE. Thank you.

Mr. McKeon.

Mr. MCKEON. Thank you, Mr. Chairman.

I tend to be kind of simple-minded and try to make things simpler rather than more complicated. I was on a school board for a number of years, and we had a lot of data on our children, and we had to furnish that to the county and to the State to receive Federal funding and State and county funding. What if we just went to the schools and asked them for the data? If we went to the State, they have to already have that information to receive the funds. Why don't we do that instead of going through this other complicated process?

In other words, they have the information on each child already. What about getting it from them instead of from the census?

Mr. BUTZ. As I understand it, Congressman, that is a possible source of part of the information that we need, but it wouldn't provide all of the information.

If we, for example, could not use Internal Revenue Service data for one reason or another in this effort, one possibility would be to use those data in combination with others. My understanding is, though, that that would not give us as complete a picture as would be obtained by using the various data sources together.

Mr. MCKEON. It may not be as complete, but by the time we wait several years to get complete data we have lost several children through the process anyway. So we could use that to go back and check. But maybe the data that they have on a daily basis to collect ADA and other sources, even though it may not be as accurate, might be more timely and possibly could be used more to achieve the desire that Chairman Sawyer is working for.

Mr. BUTZ. Could I see if Mr. Elliott, who knows more about those data than I do—would you like to add anything?

Mr. ELLIOTT. I would like to add a comment with regard to the availability of information at the school level about poverty. There is information provided by parents when children apply for the school lunch program. The question is, how consistent is that across districts and from one State to another? There is not a general measure available always in every school of poverty of parents, and we don't request that information except for the school lunch program.

So I think there would be a question with regard to information from that sort of a source unless all of us worked very hard to make sure that the information was provided on a consistent basis from school to school and from State to State as well, and it prob-

ably would require considerable supplementation of what is now done for the school lunch program.

Mr. MCKEON. Well, whatever we do, it looks like it is going to take a lot of hard work. So it is just a matter of what would be the quickest, easiest way to get it to reality.

Mr. BUTZ. Exactly. We certainly agree on that point.

Ms. RUGGLES. I think that you could use school lunch data, but I think the point that Mr. Elliott makes is a good one. We have done several studies for the Department of Agriculture on the school lunch program, and we have found that the number of parents who apply or fill out forms indicating their poverty status does depend a lot on what the lunch prices are, so that schools with relatively high lunch prices are likely to get a lot more people applying for a free school lunch than schools with relatively low lunch prices. So that kind of thing you would want to be a little bit careful about.

But you are right, it should be possible to use those data to some extent.

Mr. MCKEON. Thank you.

Chairman KILDEE. Thank you, Mr. McKeon.

Mr. Reed.

Mr. REED. Thank you, Mr. Chairman.

I would like to address a question to the panel, if you could all respond. Please correct me if I am in error, but as I understand, the poverty data is based on a national cost of living threshold. How adequate is this baseline in terms of calculating the statistics? Does this data accurately reflect regional cost of living differences? Thirteen thousand dollars a year for a family of four goes a lot further in some States than it does in my own home State of Rhode Island, and, following up on the Chairman's initial question, the good news of no poverty in Rhode Island is news that escapes the people of Rhode Island. But what is tangible is the potential decrease in assistance based on census numbers. I wonder if you could comment on that issue.

Ms. RUGGLES. I would be happy to. You are absolutely correct that the poverty line is the same poverty line across the Nation as a whole, and it doesn't include any geographic differences for price differences in different areas. We use the same, you know, $13,000 for a family of four figure in every State, regardless of what prices are in that State. It would be desirable to incorporate more State variation into the poverty thresholds, in my opinion.

I think the Bureau of Labor Statistics does have a research program underway right now to look at the question of how that could be done technically. We don't really know any definitive answer to that question so far. I think it is an area where we need more research. But you are right that there are very large cost of living differences across different parts of the country, and ideally it would be a good idea to take those into account in defining poverty.

Mr. REED. I wonder to what extent this could be disaggregated, as even within States there are great differences between certain communities and other communities, and even a statewide figure if you were at that level might mask some real differences which would deny truly impoverished families and systems the opportunity to receive assistance.

Ms. RUGGLES. You are right, States are not the only jurisdiction that you would want to look at in thinking about this question. In particular, what research that has been done on this indicates that there are big urban/rural differences and that those differences within States are often larger than across-State differences. So probably if you did come up with some sort of cost-of-living adjustment you would want to do it separately for urban and rural areas as well as separately for different regions of the country or whatever.

Mr. REED. Thank you.

Another question: In my home State, and I believe it is common in many other States, the governor is talking about consolidation of school districts and changing lines, and I wonder how, if this process of recalculation of data numbers and census numbers, if it is done on a frequent basis—how can you ensure that the problem of frequently changing school district boundaries is adequately addressed with respect to Chapter 1 distributions? Is that something that you have thought about or like to think about?

Mr. ELLIOTT. One of the things that we have done as a part of the project that the Congress mandated to the National Center for Education Statistics in 1988 is have each State give us the boundaries actually drawn on a map that can be translated into the Census Bureau tiger files that computerize all of that information. That would also have to be kept up to date if you had intercensus counts.

Mr. REED. Thank you very much.

Thank you, Mr. Chairman.

Chairman KILDEE. Mr. Gunderson.

Mr. GUNDERSON. Thank you, Mr. Chairman.

I guess I don't want to discount what any of you have said, but I sit here and I listen to you and I think of the issue, and it seems to me we start out from three basic premises, all of which are faulty. The first is that we use data that is as much as 10 years old; the second is that we use data by county, which has no relevance to the true income need of a particular school district; and, worse than that, we use all of this to determine funding for a Chapter 1 program that, once you get into the school, isn't based on income at all.

So you really have three basic premises, all of which are wrong, and we are trying to figure out how do we defend or maintain this process. It would seem to me we ought to all agree this one doesn't work and we have got to find a new one. Would anyone disagree with that?

[No response.]

Mr. GUNDERSON. Well, Mr. Chairman, now you know what your next hearing is going to be on.

Let me ask you this then. To your knowledge, is there any State that doesn't have some kind of system for determining on a regular basis the income level of people by school district?

Mr. ELLIOTT. I simply don't know. I would suspect there are a lot that do not, but I don't know. I would be glad to check that.

Mr. GUNDERSON. If you could. As Mr. McKeon said, that might be helpful.

I know in Wisconsin, and I would agree with Mr. Butz that, sure, not everybody files income taxes, but most people do, and I think that is much more accurate than this three-premised inaccuracy we operate under today.

We always, by school district, determine the amount of income and the amount of taxes you pay, and I would certainly think that that system, using that data from the previous year, would be a heck of a lot more relevant than the data we use today.

Mr. ELLIOTT. Although even so, if you want to look across the Nation as a whole—and much of this question comes because you want to have something on a consistent pattern for the Nation as a whole—information like that within Wisconsin might not be consistent at all with similar information from the State of Michigan.

Mr. GUNDERSON. In terms of income? Calculating what net income or taxable income is?

Mr. ELLIOTT. The way the calculations are made and the jurisdictions, right.

Mr. GUNDERSON. That is why we have you statisticians to figure that out.

Mr. BUTZ. In fact, Mr. Gunderson, many of the States do population estimates for pieces of geography that the Census Bureau doesn't do, and they use very different methodologies and very different data sources, and, as you are probably aware, frequently counties will even make their own population estimates for very small areas within the county using very different methods, and, as Mr. Elliott says, one of the difficulties here is how to do this in a way that is consistent across the country so that if people disagree, as they will, about the results, at least the methodology will be something that everyone will agree is fair or unfair in equal measure for all.

Mr. GUNDERSON. Let's go back to the suggestion Mr. McKeon made, again, because my Chapter 1 task force back home has suggested the very same thing, which is school lunch data, as being the most relevant, accurate data that they believe exist today in the school. Can any of you give us a reason why that would not be as accurate or more accurate than the present system?

Ms. RUGGLES. The way the school lunch program works is that forms are sent home to parents to fill out if they wish to apply for free or reduced-price school lunches. That means that it is entirely up to the parents to decide. It doesn't actually collect information on how many families are in poverty, it collects information on how many families have low incomes and want their children to receive free or reduced-price school lunches.

There is very little attempt to check that the parents have accurately filled out those forms, and there is no attempt to update it. You know, it is whatever the parents say their income is, effectively. It isn't carefully updated over time, and there is no use of records or any of that kind of thing. Nevertheless, it probably gives you a reasonable picture of the proportion of the population that is below, you know, 130 percent of the poverty line, which is the cutoff for free meals.

The biggest problem is that there are very big differences across school districts in the price of the lunches that they charge, and the work that we have done for the Department of Agriculture

really does indicate that you get a very different rate of filling out those forms and requesting school lunches in the first place if you have a lower priced school lunch than if you have a higher priced school lunch, so that there is a certain amount of inconsistency across school districts.

Mr. GUNDERSON. I am not suggesting it is perfect, but going back to one of those three faulty premises we are dealing with here, eligibility for Chapter 1 is only partially relevant to economic challenges. The other is environmental challenges. I remember making the mistake in my first term in Congress of suggesting that every student who was in Chapter 1 was financially challenged, and I had some parents very quickly take me to the woodshed on that one, and I have never forgotten that, and I think that we have to be careful here that, yes, we are going to have differences, but what is the basic purpose of the program?

Certainly in California with the base closings, I bet we are going to have some environmentally challenged children from home that has nothing to do with their present or even their former, previous years' income.

Ms. RUGGLES. Yes. I think there is no question that it would be useful to use the school lunch data at least to control, you know, as a further check on what your number of children in poverty estimates are at the school district level.

Mr. GUNDERSON. We are going to have so much controversy over the formula on Chapter 1 anyway. To what degree can any of you check to see if there is any kind of reasonably consistent income data being collected at the State level?

Mr. BUTZ. Mr. Gunderson, we are going to check on that.

I want to emphasize that in looking at these different data sources and in pointing to the difficulty, we are going to be looking for the simplest methods that we can. We are not going to start out to try to build a complex model. If we find a particular data set or several data sets that together do a very good job, we will certainly tend to use those.

The difficulty is that data of these kinds, of the sort that you suggest and others, simply haven't been tried yet at a national level, so we really don't know what is going to work, but we are going to be looking for the most efficient ways of doing it that we can. We are not going to be trying to build a complex system for its own sake.

Mr. GUNDERSON. Okay.

Thank you, Mr. Chairman.

Chairman KILDEE. Thank you.

Ms. Woolsey.

Ms. WOOLSEY. Thank you, Mr. Chairman.

I live in California, and that is, of course, a State that would really benefit from improved data, no question about it. We have a growing group of very poor citizens in my State, and we have the highest unemployment rate, so obviously people aren't doing so well. But I happen to represent an unusually affluent part of the State, which is just north of San Francisco. Even then, we have a lot of areas where there is real poverty.

Now my concern in what I am hearing here is, in my district, which is affluent, we may have to overlook and ignore these pock-

ets of poverty. Am I right in hearing this? I mean we are going to be able to identify within affluent districts and affluent counties families and schools that have need?

Ms. RUGGLES. Yes. The school district data that is currently being developed gives you the overall income picture from the decennial census at the school district level, so that to the extent that there are school districts with relatively high poverty rates in your district, it would show up in those data.

Ms. WOOLSEY. It may be one school within a district.

Ms. RUGGLES. Even then, all that means is that your district might have a lower than average poverty rate, but its poverty rate wouldn't be zero. It would still get money based on the people who were poor.

Ms. WOOLSEY. In hearing the discussion this morning, I wasn't feeling very confident about that; I'm feeling better.

I question, are we getting bogged down by using old systems and old data in what is a new world? I mean do you have available to you the newest technologies? We can have all kinds of complicated data in this day and age, and if we use new technologies it shouldn't be that complicated. We should be able to come up with what we are looking for.

Mr. BUTZ. Yes, Ma'am.

Ms. WOOLSEY. I mean you agree with me. Now, do we make this available to you?

Mr. BUTZ. We certainly try to use the most up-to-date statistical and computing and data gathering technologies. I think in some areas the Census Bureau is very up to date; in other areas, we have a ways to go. But certainly one of our principal resources that we draw on is people such as Dr. Ruggles who is an expert in the private sector, and makes herself available in a variety of ways, as do others, to help keep us up to date in terms of methodology.

Part of what we are talking about here is, in fact, using the best statistical, demographic, economic, survey, and computing technologies in the next couple of years under this legislation to try to put together a means of producing the very best numbers available. So we will certainly try to be as up-to-date as possible.

Ms. WOOLSEY. All right. Thank you.

Chairman KILDEE. Mr. Green.

Mr. GREEN. Thank you, Mr. Chairman.

To follow up on Mr. Gunderson and Ms. Woolsey, I have looked at the census data from 1990 for my own district, and we have analyzed it. There are just volumes and volumes of things I can compare in my district in an urban area of Houston that we can use, and it seems like the data is there, it is just, how do we apply it? and that is where we come in and develop the formula.

Obviously, I am not satisfied, coming from Texas, like California, that we see the current formula or method—we feel that we have just thousands and thousands of students who are not being served by Chapter 1, for example, because of the 10-year wait we have because we come from a high-growth State, particularly a border State—California, Texas, Arizona. We have a lot of growth, and so we are 10 years behind the times.

I like the idea of what Mr. Gunderson said about other methods that we can use, and Ms. Woolsey also. One of the suggestions is,

we know the poverty growth rate, for example; that can be found, and project that over the years and say, okay, if you have to go by county, although we would much prefer it to be districts, I know in Texas we do that down to the school district level, but if we have to, on Federal, base it on counties, we can do that, because we need an update.

I think the bill we are discussing, we have to have something other than every 10 years because of the changes in growth rate and the changes in what is happening. We are a mobile society. I can say during the seventies we probably had a great deal of influx from Michigan, for example, in my area, but in the eighties we were in such bad economic shape, we only looked good to folks who were making a dollar an hour, and so that is where our population growth came from. So I think we need to do something more than every 10 years, and that is the basis of the bill.

In other mechanisms, the school lunch program I am real familiar with, because that is how our compensatory education is funded in Texas, similarly Chapter 1, and it is a funding formula; we use the school lunch program for it. There are bound to be ways we can make the information more valid, I think, instead of just whatever a parent wants to say on the form, so we can utilize that if that is a better number. I don't think any of us are happy with the way the formula is provided now, whether you are from Texas or Mr. Reed from Rhode Island. So that is what we are searching for, some kind of changes.

Having looked at the data from my own district, I can pretty well tell you, although thank goodness we don't have any outdoor toilets, but I can sure tell you that the census tells me how many we have that have a bath and a half or two baths, and so we are bound to be able to tell the poverty rate and how many services we need to provide for those children through Chapter 1, and not just Chapter 1 because I know we are talking about all Federal programs, and I think that would be one of the biggest reforms we could finish here, outside of everything else, if we could provide an update on the services we provide.

My question would be, have you thought about using the poverty growth rates? I know that is updated annually or at least available with some numbers that we can verify.

Ms. RUGGLES. I think that that is essentially what this proposal is, to use poverty growth rates for each specific jurisdiction, and I think that that is what the Census Bureau is trying to figure out a way to make feasible.

Mr. GREEN. Okay. Can you do that with the numbers you have now or the information that you have now?

Mr. BUTZ. No, sir, because the poverty rates that we have now, as Ms. Ruggles mentioned in her testimony, are at an annual level and come from a relatively small survey, a survey of some 60,000 households. Once every 10 years we use the decennial census as a data source, which does provide data for very small jurisdictions.

The updating would use those decennial census data that are very detailed and would try to update them using all these other sources of information that we talked about. So it would be an updating process, but it would require both those data and the methodology that we do not now have to combine those sources of infor-

mation together to produce sufficiently accurate numbers as time marches on, every two years under the bill, at relatively small levels of geography.

Ms. RUGGLES. But basically the point of the census effort is to try and produce meaningful poverty growth rates for each jurisdiction—each county, State, town, and so on—and that is what they are using all this extra data for, is to try and put together for that specific jurisdiction how much has poverty grown or how much has it declined.

Mr. GREEN. And I also understand the concern about whether you are talking about the small numbers, the smaller jurisdictions. Of course, I think our goal on the committee is to serve actually the children instead of the jurisdictions, because that is where the funding goes to. Hopefully the Chapter 1 benefits that child instead of that district, or that county, or even that State.

Thank you.

Thank you, Mr. Chairman.

Chairman KILDEE. Mr. Becerra.

Mr. BECERRA. Thank you, Mr. Chairman.

I apologize for being late and not being able to listen to all of the testimony, and perhaps these questions have been answered, but let me ask a little bit about the undercount that took place in the 1990 census and what the impact of that undercount will be with any type of calculations and formulas we come up with.

As I understand it, for example in California, I understand that there was an undercount in the area of over a million people, and I suspect that we will find that, when you take a look at the poverty population, that would be even a higher percentage of the undercount. How are we going to try to tackle that undercount when we come up with a formula?

Mr. BUTZ. The decision of the Secretary of Commerce was not to adjust the 1990 census for differential undercount, and the director of the Census Bureau made a subsequent decision not to adjust the population estimates for undercount, and it is those estimates that I described earlier as one of the foundations of the population estimates, which in turn is one of the foundations for the poverty estimates that we are discussing here.

So under those decisions, those differential undercounts, as measured by the Census Bureau, will be carried on through the decade until the next census occurs.

Mr. BECERRA. Is there any way for the Department of Commerce to reexamine its decision not to incorporate those undercounts?

Mr. BUTZ. I am afraid answering that is beyond my knowledge and my authority.

Mr. BECERRA. Well, let me ask it this way. Are you aware of anything that prohibits the Department from reexamining that undercount?

Mr. BUTZ. I am not aware of anything.

Mr. BECERRA. Okay.

Again, I think the next two questions will be in line with what I asked at the beginning. Amnesty applicants under IRCA will all be eligible to become citizens over the next year or two. They are, for the most part, a population that is somewhat mobile. Is there any

fear that we may have missed some of these folks or that they are a good portion of the undercount?

Mr. BUTZ. That is a population that has a relatively high differential undercount, along with some others, but to the extent that these people are picked up in administrative record systems in the school lunch program and the AFDC program of the sort that we have suggested, and to the extent that that information is modeled or put together with the baseline information from the 1990 census, there will tend over time to be some accounting for those people in the combined estimates. How that will work and how much of the differential undercount could be made up that way I don't know.

Mr. BECERRA. Does it help that the Immigration and Naturalization Service has documentation for each of these individuals and, I suspect, family members as well? Can the Census Bureau use the information that is housed within INS?

Mr. BUTZ. Yes, sir, we do use that information as a means of developing our estimates of the numbers of undocumented aliens in the U.S., and we also use that information to some extent in coming up with our estimates of the numbers of people missed in the census.

Mr. BECERRA. The final question I have relates to the migrant population, which of course is probably the most undercounted population. I looked through your testimony, and you mentioned that the small areas are those areas that have the largest undercount. Are we going to try to do anything to try to take care of the undercount that may have occurred with the migrant population?

Mr. BUTZ. Well, first, I didn't mean to state or imply that the small geographic areas necessarily have the largest undercount. They tend to have the largest sampling error in producing estimates.

For the 2000 census, we are, with the excellent cooperation of Chairman Sawyer and his leadership, making enormous efforts to try to find ways to reduce the differential undercount for the 2000 census. It is a daunting job, but we are doing our best to plan ways of doing it.

Mr. BECERRA. Any chance of accelerating that for your estimates coming up in the next couple of years?

Mr. BUTZ. It is really a job that involves the census much more than it does the estimates. The census is the basis for the estimates, and to the extent that the census is right, the estimates will tend to be right.

Mr. BECERRA. Mr. Butz, I am a bit apprehensive about what you say, because it is clear that we are using figures that may lead to some erroneous results in terms of our estimates. Is there anything you see that we can do short of waiting until the year 2000 to try to correct some of the inaccuracies in the estimates for some of the populations that might be missed or more mobile?

Mr. BUTZ. There are methods available for adjusting the estimates process, for adjusting the census as a base for the estimates, but the decision has been made that below the State level, certainly, the accuracy of those adjusted estimates is unclear and cannot be demonstrated to be better than the accuracy of the estimates

based directly on the census without adjustment. That was the decision that was made.

As I say, there are methods of doing it. The question is, do those methods produce demonstrably more accurate data or not? The decision has been made, no.

Mr. BECERRA. Thank you.

Thank you, Mr. Chairman.

Chairman KILDEE. Mr. Payne.

Mr. PAYNE. Thank you very much, Mr. Chairman.

I am very sorry to be this late, and it is good to see Congressman Sawyer still on the job, looking over the census. We spent a lot of time trying to see that the 1990 census would be done correctly, as I wished that the 1980 census would have been done correctly, and as we agonized with the 1970 census which was done poorly, and I decided to do some research and found out that when they did the first census it was reported that it was inaccurate back, I think, in 1890.

Mr. BUTZ. Yes, sir. The President was upset about it at that time.

Mr. PAYNE. And so the thing that really amazes me is the fact that the undercount is accurate. We have a way of knowing how many people were not counted, and that always amazes me because we simply can't use that because it is circumstantial evidence. We have to really see the body or see the note from that individual.

So I really would hope that there could be other approaches. They tell me that they do have ways of knowing deaths and births and utility bills and things of that nature and that a fairly accurate accounting is being kept, but the fact that the enumerators or the individuals have not responded, therefore they are ineligible, this becomes serious.

For example, in New Jersey about $33 million is missing because of Title I. There is a total of 10 programs in Title I of the Elementary and Secondary School Act which rely on population figures and formulas for determining allocations. Programs under Titles II, III, IV, and V rely on population for determining allocation standards, and we also know the whole question of Chapter 1, and it is really creating a very serious hardship in my district where schools are seriously being recommended they downsize their Chapter 1 programs.

I missed the testimony, and I am sure that we will attempt to have a more accurate census in the future, but the continued undercount I think is totally unfair, especially in light of the fact that it can be done in a more judicious manner.

Also, I get confused by the judge's rulings where he will rule that—is it Judge Green who has been dealing with this?—the judge in the recent case, where he acknowledged that there was an undercount, but would use it partially for one reason but wouldn't use it for another. I think that that kind of confusion even from the judiciary where we can't get a clear picture—it is hard for both people to be right, and the judge is saying, "You are both right, or either we are both wrong," and someone has got to be right, and someone has got to be wrong. I mean that is the way they taught me in school.

So I just wonder what type of resolution, because it does do irreparable damage, particularly to the poor districts, and I just wonder

if anyone has any quick, one-minute response, because I am sure you handled it in your testimony.

Mr. BUTZ. Congressman, the principal goal of planning for the 2000 census is to reduce the differential undercount. That is our principal objective. It is a very difficult problem. You have said a lot there, and it is hard to know how to respond.

I would say in terms of the difference of interpretation of numbers that we feel that we have a very good handle at the national level of what this undercount is, how much it is, how it is distributed across different groups—renters, owners, by racial and ethnic groups. The difficulty is trying to work that down to very small geographic levels, and in that sense it is somewhat analogous to the problems that we are talking about here of producing data for school districts. Again, it is that low level of geography that presents the problem.

For example, you mentioned birth and death records. At the national level, we do go back, way back, decades back, and add up how many people were born and died, moved in and moved out, but that is national data, and where someone was born isn't where they live now.

So it is very difficult, and, unfortunately, it is subject to different interpretations as to whether the data are useful for one purpose and not for another, and I share a good deal of your frustration.

Mr. PAYNE. Finally, also you find that there becomes a differential within a State geographic area. For example, several of the cities, I encouraged that they sue the Census Bureau because of the fact that we knew there was an undercount. We had an undercount in 1980, and we have restored about 30,000 people, almost 10 percent of the population in the City of Newark, back when the 1980 census was done, and we also had some numbers put back in 1990. But the undercount numbers which affected my district, if they ran them through the States throughout the Union, the undercount in New Jersey was less than it was projected for some other States. Therefore, New Jersey would have lost under the recognition of the undercount.

So the State takes a position, the governor's position—who is a former Democratic colleague here in the House—took a different position than I took as a representative of Newark and Elizabeth and many impoverished cities because it would have benefited our district, but overall the State, in comparison with the national situation, would have lost.

So we do find also that the integrity is questioned when State officials do what they feel is best for the State and, once again, the poor districts suffer by virtue of the decision which would be considered almost local from a national standpoint on a State level, but when we get down to the difficult real local areas, it makes it extremely more difficult.

Thank you very much.

Chairman KILDEE. It is my understanding that Caesar Augustus questioned the figures made during the census when Sabines was governor of Syria. So there are always questions on the census, and I am sure you are trying to upgrade your systems ever since the time of Caesar Augustus to try to get more accurate figures. I know

you are working hard at that, and I know Mr. Sawyer is very interested in accomplishing that.

One question: Commissioner Elliott, you stated that in the absence of school district updates from the Census Department, county updates of the poverty counts would be useful for updating the distribution of Chapter 1.

Mr. ELLIOTT. Right.

Chairman KILDEE. Are there any special problems for producing district level updates for school districts that cross county boundaries?

I ask the question because I represent basically three counties in Michigan. One of the counties is considered the first or second most affluent county in the country, and the other one is the county in which the city about which Michael Moore wrote the book "Roger And Me" is located. One of the school districts crosses that county line, one in the more affluent county and one in the county that has very high unemployment. How would we be able to ascertain the figures there?

Mr. ELLIOTT. There would have to be a similar sort of thing that is done for the mapping of information from the 1990 census in each school district, and that is, where the boundaries occur you try to draw a line where the actual boundary is and then proportion the population on each side of the line. That frequently takes a lengthy conversation within a State to agree on what will be the basis for the proportion that is split, and that is passed along to us, and then we give it to the Bureau of the Census so the information can be produced back by that new boundary. But the same sort of thing would have to happen there where something went beyond the jurisdiction of a single county to pick up a portion of each of the two counties in which the district was located.

Chairman KILDEE. And under the present situation, it is the State that makes that determination, how much money would flow to that county, and then also the State makes the determination of how much money would flow to a particular school district within that——

Mr. ELLIOTT. Within the county, and then if more than one county is involved, then they would have a larger jurisdiction and fit within the combination.

Chairman KILDEE. Thank you very much.

Mr. Sawyer.

Mr. SAWYER. Thank you, Mr. Chairman.

Just an observation. It has been enormously impressive, the amount of effort that the members of this committee have made to come to grips with the substance and the implications of the subject that we are discussing today. I am grateful for the questions that have been asked, frankly, on both sides of the table.

Let me ask a question, and I direct it primarily at Mr. Elliott and Ms. Ruggles. We are trying to find, in addition to the way in which we can statistically improve the quality of intercensal estimates, the kind of administrative data that will give us the best indications of how well we are doing and where we ought to make those adjustments.

Can you talk about the kinds of data, the sets of data, that have the greatest consistency from one State to another? We can talk

about the failures of certain kinds of data, lunch programs and others, but can you talk about the data that provides the greatest consistency?

Ms. RUGGLES. I think probably the food stamp program for the low end of the population is the data set that you would want to use, because it is the only program for low-income people that has the same rules and the same income cutoffs and so on, no matter where you live.

For example, the Aid to Families with Dependent Children program has different benefit levels in different States, and that affects the eligibility, so it isn't very consistent across States. But the food stamp program is pretty consistent, and we do have a reasonable set of case records for that program.

Mr. ELLIOTT. I think probably Ms. Ruggles is more familiar with those data outside of the education system than I am. Actually, I would welcome a challenge within the National Center for Education Statistics to work with our State Cooperative Statistics Program to see whether either there are some available data sets that might be more consistent or, on the other hand, whether some of the ones around, like school lunch, might be improved considerably. That is simply not something that we have explored, and I think that is something that could be done.

Mr. SAWYER. Just with regard to the school lunch program, the poverty thresholds are at 130 percent or something like that. It would be fine if the distribution of poverty between those two were consistent across the country, but I know that when I think of Mr. Gunderson's State, there are fewer people who are impoverished, but when they are impoverished they are extremely impoverished by comparison to many other places. That kind of consistent, smooth flow just doesn't exist.

It brings me to a second question, and that is the notion of geographic cost-of-living differences. We have had several members who have talked about that in one way or another. Is there a reliable way to build those kinds of considerations into our estimates?

Mr. BUTZ. Well, the responsibility and the expertise for that within the Federal Government is principally in the Department of Labor at the Bureau of Labor Statistics, and, as Pat Ruggles mentioned earlier, they are and have in the past been concerned about this. It is apparently a severe problem, a data problem, in finding the information that will reliably produce better cost-of-living measures at local areas than if one simply uses the national average overall. But I believe they are working on it still.

Ms. RUGGLES. Yes, they are. They do have a research program in this area. In addition, I think the Poverty Panel that the National Academy of Sciences has going on right now will probably come up with some recommendations in this area.

Mr. SAWYER. Let's see. Maybe by September of 1994, right?

Ms. RUGGLES. As a matter of fact.

So I think that there is work going on in this area, but it is a difficult technical problem. What constitutes a representative sample of prices for particular places is a pretty tough thing to figure out and to compare from place to place.

Mr. SAWYER. In closing, Mr. Chairman, all of that really goes back to what you started us with, the Mollie Orshansky numbers.

What we are trying to deal with here is to come together with a three-legged stool that works. We are trying to get an accuracy of measurement that comes from timeliness to work together with a definition of poverty that may well have changed substantially since the mid-1960s—end of the sixties, when it was made official, and a distribution formula that recognizes both in a way that produces the results that we want. In any case, it seems to me that accurate measurements are the cornerstone of that kind of sound policy.

Thank you very much.

Chairman KILDEE. On that point, I know in my district it would be hard to take just AFDC figures to determine poverty because we have so many people in my district now who are the working poor, and they are in great need. So there has to be a constant reevaluation of what constitutes poverty and how then we distribute funds, all types of funds, including Chapter 1 funds, based upon the poverty.

I lived in the City of Flint when you could quit school on Tuesday and go to work for Buick on Wednesday, and those days are gone forever. We find so many people, and we are sometimes assured, "Don't worry, we'll move into a service industry economy," and we are moving into a service industry economy in Flint, and people are very poor in doing that.

So I think we have to constantly upgrade the methods of counting, upgrade the methods of determining various things about people, and constantly upgrade our formula to distribute the dollars based upon those figures.

Before we go to the next question, I would like to acknowledge the presence of a very dear friend and former chairman of this committee, whose picture is right behind me, but there he is right in front of me over there, Mr. Gus Hawkins.

[Applause.]

Ms. Woolsey.

Ms. WOOLSEY. No questions, Mr. Chairman.

Chairman KILDEE. Mr. Payne.

Mr. PAYNE. No, thank you, Mr. Chairman.

Chairman KILDEE. Without any further questions then, you have been very, very helpful. Certainly you are the professionals, the experts, the parite, as we use a Latin term, in your field, and even very, very helpful to this committee. I want to thank all of you. We look forward to continuing to work with you as we work on H.R. 6, the Elementary and Secondary Education Act, and on Mr. Sawyer's bill, H.R. 1645. The record will remain open for two additional weeks for any additional testimony.

Thank you very much.

[Whereupon, at 11:52 a.m., the subcommittees were adjourned.]

[Additional material submitted for the record follows.]

Statement of Hon. Donald M. Payne, a Representative in Congress from the State of New Jersey

Chairman Kildee and Chairman Sawyer, I would like to take this opportunity to thank both of you for bringing these distinguished panelists before us today to discuss H.R. 6 and H.R. 1645.

I am honored to be a part of this important process that will examine whether accurate data is being provided in between census years. Census data is extremely important because it often serves as a basis for legislation, case studies, and research models, among other things. The proposed legislation would require the Census Bureau to produce poverty estimates for States, counties, cities and school districts.

Currently, data is only collected during the decennial census and frequently does not reflect the overall poverty situation in a given area.

Census data is particularly important in light of reauthorization under Chapter I. Education allocation formulas usually rely on population groups that approximate the intended beneficiaries, and for Local Education Agencies, the benefit is intended for educationally disadvantaged children living in relatively low-income areas. Furthermore, for other ESEA allocation formulas, a variety of populations are used for the calculation of grants.

A total of 10 programs under Title I of ESEA, rely on population formulas for determining allocation. Four other programs under Title II, III, IV, and V—also rely on population for determining allocation standards.

In poor, inner city districts, the availability of accurate, valid data is crucial in improving the lives and education of children. I strongly believe that we need data that will reveal the true poverty situation in our country. I encourage a full and engaging discussion on this legislation.

I would like to thank all of today's panelists for taking the time to present their testimony for us today. I look forward to hearing their comments.

Statement of Hon. Don Edwards, a Representative in Congress from the State of California

I would first like to thank Congressman Tom Sawyer, Chairman of the Post Office and Civil Service Subcommittee on Census, Statistics, and Postal Personnel as well as Congressman Kildee, Chairman of the Education and Labor Subcommittee on Elementary, Secondary, and Vocational Education for giving me the opportunity to share with you the Delegation's views on H.R. 1645, "The Poverty Data Improvement Act of 1993."

H.R. 1645 is a critically important bill which would treat all the people of the United States equally. It would require the Census Bureau to produce and publish poverty estimates for States, counties, cities, and school districts every two years beginning in 1995. This would replace the current system which utilizes decennial census poverty data to annually allocate the estimated $20 billion in grants to State and local entities. The reliability and usefulness of this data, collected once every 10 years, quickly becomes outdated and ineffective in accurately allocating the funds to the places and people of greatest need.

Californians are particularly aware of how the accuracy of census and poverty estimates affect the State's ability to get its fair share of Federal dollars. The unique characteristics of California's population make it among the most difficult to count accurately. Based on research completed at the Population Resource Center, a nonprofit, non-partisan organization, California will gain 30 percent more people and absorb three-eighths of the population growth nationwide between the years 1990–2005. This increase includes the integration of nearly one-half of all legal foreign immigrants and refugees into the United States annually. The spiraling growth of California places an incredible burden on the resources of our State and local governments which the data from a decennial census cannot accurately gauge or accommodate.

While most areas of the country are starting to feel the benefits of an economic upturn, California still finds itself reeling from the recession. With numerous base closures scheduled, many defense plants shutting down, and statewide budgetary problems, the crisis will likely worsen. However, this is not a question of pitting one State, or region, against another. The question before the Congress today with this legislation is fairness. We have an obligation to use the numbers that most accurately reflect the needs of our State and local governments. That was Congress' intent when it passed legislation requiring the allocation of funds based on population data. Everyone "wins" by having the most up-to-date data to distribute the Federal grants to State and local governments.

One of the most prominent examples of the need for this legislation is the Chapter 1 program. For current 1992–1993 school year, Chapter 1 funds were allocated using the 1980 census. This census data was collected in 1979, making the income data 14 years old and hopelessly out-of-date. In California alone, using 1980 census data instead of 1990 data for just the 1992–1993 school year, this translates into a loss of over $125 million. This is magnified manyfold over the course of the 14 years, not just in funding but also in the loss of services and goods which such moneys could provide for the underprivileged children who need it the most. H.R. 1645 would resolve this disparity with timely poverty data updates below the national level.

I understand the Census Bureau has already prepared a proposal for research and development of the methodology for producing the data which H.R. 1645 would require. The cost of this program is set at only $450,000 annually. This figure is only a fraction of the amount of funds that are allocated each year on the basis of poverty data. In such times of scarce funds and budget reductions, this clearly is an effective and efficient use of taxpayer money.

Last August, I had the opportunity to testify before the former Director of the Census Bureau, Dr. Barbara Everitt Bryant, in favor of adjusting the intercensal population estimates. Again the issue was fairness. Unfortunately the Bureau ultimately decided not to use the Post Enumeration Survey for its annual census adjustment. I was disappointed in their decision. H.R. 1645 goes a long way toward mitigating the problems debated at that time.

Mr. Chairman, during these trying economic times and budgetary cutbacks, we must allocate what limited resources we have in the most effective, efficient, and fair manner. Without accurate periodic poverty estimates, many of our most desperate and destitute people will find their needs going unmet. H.R. 1645, "The Poverty Data Improvement Act of 1993," will provide an inexpensive, effective way of fairly distributing this shrinking source of funds. I will be working diligently for its passage and I urge my colleagues to join me in strong support of this legislation.

STATEMENT OF GAIL IMOBERSTEG, DIRECTOR, FEDERAL LIAISON OFFICE, CALIFORNIA DEPARTMENT OF EDUCATION

The reliance of poverty estimates generated through the decennial census for the allocation of Federal program funds presents some important problems with regard to the allocation of education funding. Due to the rapidly changing demographics in the United States, the data used to distribute funds for programs such as Chapter 1, Hawkins-Stafford Elementary and Secondary School Improvement Act [Public Law 100–297], are hopelessly outdated before they are used. For example, the allocation for fiscal year 1992 program funds for the Chapter 1 program was based on data collected for the 1980 census. Since this data actually incorporates 1979 poverty estimates, the data was actually 14 years old when it was used to distribute funds. The 1990 poverty data will be five years old by the time it is used to allocate Chapter 1 funds.

The use of data 5 to 14 years old to direct Federal funds raises some significant issues of accuracy and fairness. Between 1980 and 1990, California became home to an additional 247,162 poor children. This represents a 38 percent increase! These poor children, many of whom are limited-English proficient, have now been in California classrooms for years. Since they are uncounted in the Chapter 1 formula, there have been no additional funds provided to serve these children under this program. With diminishing fiscal resources, the targeting of funds to the intended recipients is particularly important. The use of outdated census data for the distribution of Chapter 1 and other program funds flies in the face of efforts to target resources fairly.

Since the Chapter 1 program formula is used to allocate funds for a number of other education programs, the accuracy of this data has had far-reaching implications for education programs in California and throughout the Nation. In addition, the enormity of demographic changes that result every 10 years exacerbates conflicts among States that are struggling to provide programs with diminishing State and Federal resources. If poverty estimates were provided every year, or at least every two years, the distribution of Chapter 1 funds would follow the children as populations shift throughout the decade. This would not only enhance the targeting of funds, it would also eliminate the enormous shift in funds that occur after each decennial census and allow for the gradual absorption of decreases and increases of funds as populations shift.

The California Department of Education is strongly supportive of Congressman Thomas Sawyer's bill, H.R. 1645, which requires the Census Bureau to produce poverty estimates every two years for States, counties, cities, and school districts.

ANSWERS TO QUESTIONS SUBMITTED BY THE CENSUS BUREAU, DEPARTMENT OF COMMERCE, WASHINGTON DC

First set

Question 1a

You expressed concern about access to IRS data for the proposed intercensal small area poverty estimates program. Would the direct requirement for these estimates in H.R. 1645 and direct funding, as provided in the House version of the fiscal year 1994 Commerce, Justice, State and Judiciary Appropriations bill, eliminate that concern?

Answer 1a

We met recently with IRS officials. They expressed their support for this important project and assured us that we will have access to the data we will need to carry it out.

Question 1b

Does [or would] the Bureau have access to W-2 and 1099 forms, if those records become necessary for the poverty estimates program?

Answer 1b

Based on our recent meeting with the IRS, we believe that we will have access to whatever records are necessary for the program. While all details regarding the level of access have not been worked out, we have agreed to work together to develop the best possible estimation methodology.

Question 1c

Will you need additional legislative language to facilitate the transfer of data from the IRS to the Census Bureau?

Answer 1c

No. We believe current statutory authority is sufficient.

Question 2

In addition to separate tabulations for the number of poor school-age children, will the Bureau be able to report separate tabulations for other age groups in poverty [such as "over 65 years of age"]?

Answer 2

As you know we have developed a plan to provide updates of the 1990 census income and poverty estimates during this decade. This plan would be funded by a consortium of departments and agencies who use these data for the allocation of funds and for other administrative purposes. The plan, as it now stands, is aimed at meeting the needs of this consortium. Since none of the members have requested poverty estimates for population subgroups other than children, we have not planned to make estimates for any other demographic breakdowns. Expanding the scope of the project to include estimates for other population groups may be possible. This would, of course, require additional work in both the population estimation and income and poverty estimation.

Second set

Question 1

What would be the most reliable and cost-effective way to update census data on children in poor families, whether at the county level or the LEA level, between now and the next decennial census, so that we can minimize the likelihood of large population shifts, and attendant Chapter 1 allocation changes in the future?

Answer 1

We have developed a proposal to provide updates of 1990 census income and poverty estimates during the decade. Since we have never attempted to make these kinds of estimates, it is difficult to say what would be the most reliable and cost-effective ways to produce these data. The first phase of our proposal is to conduct the research that would answer that question.

The large population shifts that you reference reflect real changes in the size and geographic distribution of the population during the 1980s. A program to provide periodic updates of the 1990 census poverty estimates during this decade would reduce the time period between poverty estimates from 10 years to an interval of 2 to 3 years. These estimates could be used to make more frequent adjustments to the allocation of Chapter 1 funds. These more frequent adjustments would be smaller, but, their net effect over the 10-year period between censuses would reflect the same overall change as measured by the censuses.

Question 2

There is legislation that would require changing county data on poor school-age children by the statewide rate of increase or decrease in all school-age children. What is your opinion of this proposal? Is its implicit assumption—that local trends in school-age poor children closely track statewide changes in total school-age population—valid? Are there States where during the 1980s, total and poor school-age populations moved in opposite directions?

Answer 2

We are unable to answer the first two parts of the question at this time. Since we have not received a copy of the proposal, it is not possible for us to comment on it at this time. With regard to the second part of the question, we have not made any analysis of the relationship between local trends in poverty and statewide changes in the size of the total school-age population. We do have data to answer the third part of the question. These data show that during the period between 1980 and 1990, 14 States experienced increases in the number of poor school-age children [based on related children age 5 to 17 years old] while experiencing declines in the total population of such children. We found no States in which the number of poor children declined as the total number of children increased. There were 12 States in which both the number of poor children and the total number of children increased. In the remaining 25 States [including the District of Columbia] both the number of poor children and the total number of children declined between 1980 and 1990.

Question 3

What would be the advantages and disadvantages of conducting a mid-decade census in 1995?

Answer 3

The clear advantage to conducting a census [including collection of sample data on income and other measures of poverty] in 1995 would be to produce updated statistics five years earlier than would occur by waiting for the next decennial census. Thus, data would be available in 1996–1997 rather than 2001–2002.

The biggest disadvantage would be cost. It likely would cost well over $1 billion to conduct an actual census [as opposed to a large sample survey] that could produce data [both 100 percent and sample] for small areas [such as school districts or counties] similar to that for the 1990 census. Even if funds were available beginning in fiscal year 1994, it is unlikely that all the preparations [particularly address list development] needed to take a census could be completed in time to conduct a census the following year.

For Witnesses from the Census Bureau and the National Center for Education Statistics

Question 1

What is the current status of the proposal that has been worked on by Census Bureau staff to provide updated county-level estimates of children in poor families biannually beginning in 1995? Does the Department of Education intend to support and contribute to this project? What are current estimates of its cost, and the probable reliability of the population estimates it would produce?

Question 2

Do you plan to keep up-to-date the LEA maps prepared for the purpose of compiling the 1990 census by LEA?

Answer to 1 and 2

The two questions assigned to the Census Bureau and the National Center for Education Statistics witnesses are largely within the scope of NCES programs. We, therefore, defer the answering of these question to them.

NORTHEAST-MIDWEST CONGRESSIONAL COALITION

Testimony
of

Hon. Dean A. Gallo

Co-chair, Northeast-Midwest Congressional Coalition

before

Subcommittee on Census, Statistics and Postal Personnel
House Post Office and Civil Service Committee

and

Subcommittee on Elementary, Secondary and Vocational Education
House Education and Labor Committee

July 13, 1993

U.S. HOUSE OF REPRESENTATIVES
530 Ford House Office Building, Washington, D.C. 20515 • (202) 226-3920

Messrs. Chairmen, I am testifying today on behalf of the Northeast-Midwest Congressional Coalition, which is co-chaired by myself and Rep. Marcy Kaptur and which is honored to have both of you as active members. I am pleased to present for the record some comments on the adequacy of the nation's poverty statistics and the Coalition's suggestions for how they can be improved.

To begin, let me congratulate the panel on two counts: first, for the attention you are paying to this issue, and second, for what a joint hearing says about both the legislative role in setting statistical policies and the user considerations that should go into the making of those policies.

First, you deserve credit and the thanks of Congress for taking seriously a subject that some might dismiss as too arcane and esoteric for elected officials to bother about. You, in contrast, recognize that, as legislators, our understanding of social conditions and our ability to formulate responsive programs depend on our having good information. But, our dependence on statistics goes far beyond these policy development uses. Indeed, because of congressional decisions, taken over the past twenty-five years, to use number-driven formulas for allocating federal funds and defining eligibility for benefits, we are now in a position where the implementation of the vast bulk of federal domestic policy rests squarely on the nation's statistical programs. Statistics drive billions of dollars this way and that--into some hamlets, into some cities, towards some persons, away from others. When statistical data programs (1) fail to measure the conditions we policymakers think they are measuring, or (2) fail to measure them in a timely fashion, then the policies we enact also will fail. The foremost example of these two types of statistical failures--not measuring what we intend and not doing it promptly--is poverty data, the subject of this hearing.

Before turning to that subject, however, I want to explain my second reason for congratulating the panel. It is significant that the Subcommittee on Census, Statistics and Postal Personnel--the authorizing body for the Census Bureau's statistical programs--and the Subcommittee on Elementary, Secondary and Vocational Education--which depends on census data for the implementation of its policies--have chosen to hold a joint hearing. That says something important about the making of statistical policy. It says that statistical programs cannot be designed or evaluated in isolation from the ways the data are used. The poverty data are a case in point; defenders often excuse the shortcomings of these data by noting that they were not designed for the way Congress now directs that they be used in allocating funds and defining who is eligible for benefits. Somehow, these defenders manage to imply that the fault is not with the data but with Congress for using data the "wrong" way. I would submit that this formulation needs to be reversed: BECAUSE OF THE WAY CONGRESS USES THEM, THE POVERTY DATA NEED TO BE REDESIGNED TO DO WHAT CONGRESS INTENDS.

By holding a joint hearing, you are doing something very significant here today. You are saying, in effect: (1) that statistics need to be designed with the users in mind; (2) that Congress itself is a primary user; and that (3) the implementation of federal domestic policy is now the first and foremost use of the products of the federal statistical system. Elected officials want a statistical system that anticipates and responds to the data needs of the policymaking and policy implementing process, a system that grows with the way data are used. Unfortunately, the statistical system has not been responsive to the way Congress is using poverty data.

I want to concentrate today on two ways in which the poverty data fail to meet the needs of Congress--their lack of timeliness and their lack of validity. The timeliness problem is obvious; using data that are ten or more years old to carry out federal domestic policy--as we have done with Chapter 1, the Job Training Partnership Act, and other programs--is irresponsible. The Chairman's actions to develop legislation on the subject of more timely poverty numbers are commendable. The Census Bureau's initiative to produce biennial updates of poverty data is long overdue.

The timeliness problem thus seems well on its way to solution. So, I want to call the panel's attention to an even greater problem with the poverty statistics--that of their validity. Simply stated, while we assume that the poverty data provide counts of the number of persons in need, in fact they do no such thing. The poverty counts are numbers below one national income standard that does not take into account the very real cost-of-living differences from one area of the country to another. Because of cost-of-living differences, a person whose income is just above the poverty level but who lives in a high cost-of-living state such as Connecticut or California can be much worse off than a person whose income is just below the poverty line but who lives in a low cost-of-living area such as Louisiana or New Mexico. Despite greater need, the former is not counted as poor; the latter is.

Our current measurement of poverty, in short, does not measure relative need. And yet, relative need is the very concept Congress is groping for when we design allocation formulas for programs. Relative need is what we <u>think</u> the poverty data is measuring. Not so. By using these data, we misdirect funds, sending proportionately more funding relative to need to some areas and less to others. This means that funds adequate-to-need for remedial education, for summer jobs, for training, and for other human development purposes, are systematically denied to places like Newark, Detroit, Cleveland, Los Angeles, the former mining towns of western Pennsylvania and northern Minnesota, and remote rural locations in Vermont and Iowa. Needy persons who by any commonsense consideration are "poor" continue to go without. The resulting patchwork of funding priorities amounts to no federal policy at all.

Addressing the timeliness problem only deals with half the issue. It is equally important, perhaps even <u>more</u> important, that, in reforming the poverty data, you deal with the present inability of these data to measure relative need. The way to do this would be to use a state-by-state cost-of-living index to produce state-specific poverty income standards. A state's residents whose income falls below the state-specific poverty standard would then be counted as "poor." The poor from all states, added together, would constitute the poverty population of the nation.

I would like to submit with my testimony two state-specific cost-of-living indices to show what can and should be done to make the poverty data meaningful. You will notice that both researchers, using somewhat different methodologies, arrive at similar results in terms of the rank order of states and the magnitude of cost differences separating them. Differences of over 50 percent in the cost of living distinguish one state from another. With differences this wide, it is clear that our present use of one national income standard to define poverty does grave injustice to the concept of need. The procedure I am advocating would multiply the national poverty standard by the values in a cost-of-living index to create poverty standards for each state. The second attachment to my testimony demonstrates this procedure.

I want to anticipate two doubts that may be raised about the idea of state-specific poverty standards. First, some analysts will posit that cost-of-living differences within states may be as great as differences between states--the example of New York City and upstate New York is often cited in this regard. Although the cost-of-living does vary within a state, it still is the case that a state-cost-of-living-adjusted poverty standard will come closer to the truth of what level of income constitutes need in any one state than does our present national poverty standard. In addition, the seminal work by Dr. Walter McMahon, Professor of Economics at the University of Illinois, shows that while there is a 57 percent difference in purchasing power between the highest and lowest states, the variation within states is in the lower ranges of 22 to 35 percent. Finally, the poverty data we have now are arbitrary and manifestly incorrect. What I am proposing is an obvious improvement, which may itself be improved further. In short, I want to say to the panel: don't do nothing just because you may not be able to achieve perfection.

Second, some may argue that we need more research on this idea. In answer, I would say that while research on state cost-of-living differences is not voluminous, it is respected. Indeed, the U.S. Department of Education has arranged to have one of these indices updated this fall when new census data become available. Instead of ordering more research or calling for more study, I would urge the panel to move forward and direct the Census Bureau to implement the idea of state poverty standards. I know, from the Northeast-Midwest Coalition's experience with legislation on another data set (unemployment insurance [UI] wage records), that nothing spawns research on a subject or leads to technical breakthroughs faster than Congress directing an agency to implement something.

All of us here have great respect for the research and analytical capability of our statistical agencies. We as policymakers don't have to figure out all the details of how to design data. We do need to set the direction, to point out the goal, and to issue the marching orders. I am confident that, if Congress enacts legislation directing the development and employment of state-specific poverty standards, our statistical agencies will respond to the challenge quickly and competently.

Finally, let me offer you an additional reason for moving on the issue of state poverty standards now. As you well know, it is not often that the attention of members of Congress is captured by an issue in the realm of statistical policy. Right now, the effect of the 1990 Census data on allocations for our remedial education program, Chapter 1, and the pending reauthorization of that program are making members more than usually sensitive to the importance of statistical policy and the need for attention to the statistical system. I fear that if you do only half the job now--if you enact what the Census Bureau has already indicated it is prepared to do, i.e. produce more frequent poverty data--that you will never again have the opportunity to address the second part of the problem--the ability validly to measure need.

Again, the experience of the Northeast-Midwest Coalition's initiative on UI wage records is instructive. If we had treated the issue piece by piece with one bill at a time, we would never have been able to sustain the effort. I urge you to capitalize on the fact that Chapter 1 is creating a natural constituency for true, comprehensive reform of the poverty data that addresses the two problems of timeliness and validity. You will not have this chance again for a long time.

State Rankings in Cost-of-Living Studies

McMahon Study, 1988				Nelson Study, 1989		
State	Rank	Index[1] Value	Percent Change 1977-1988	State	Rank	Index[1] Value
District of Columbia	1	124.9	19.4	Alaska[2]	1	
Connecticut	2	123.7	2.9	Hawaii[2]	1	
New Jersey	3	119.1	2.1	Connecticut	2	125.8
Massachusetts	4	114.0	5.8	New Jersey	2	125.8
Hawaii	5	113.9	n/a	Massachusetts	3	121.8
New York	6	110.7	0.3	District of Columbia	4	121.0
California	7	110.2	2.2	New York	5	113.2
Maryland	8	109.4	-3.4	Maryland	6	110.8
Illinois	9	107.7	4.5	New Hampshire	7	107.9
Minnesota	10	104.7	3.8	Rhode Island	8	106.8
Iowa	11	102.5	7.2	California	9	105.9
Michigan	12	102.2	1.5	Delaware	10	103.2
New Hampshire	13	101.9	-4.4	Pennsylvania	11	101.3
Delaware	14	101.7	-8.5	Arizona	12	100.8
Alaska	15	101.7	n/a	Colorado	13	99.8
Colorado	16	101.6	1.0	Nevada	14	98.1
Washington	17	101.5	1.8	Illinois	15	97.4
Rhode Island	18	101.3	-2.2	Florida	16	97.3
Virginia	19	101.2	7.9	Virginia	17	97.1
Wisconsin	20	101.1	1.4	Washington	18	97.0
Ohio	21	100.7	0.8	Ohio	19	95.8
Pennsylvania	22	100.3	5.4	Minnesota	20	95.6
Nebraska	22	100.3	5.2	Michigan	21	95.4
Oregon	23	99.5	1.4	Wisconsin	21	95.4
Kansas	24	98.0	4.5	Oregon	22	94.2
Nevada	25	97.1	-9.1	Texas	23	94.1
Missouri	26	96.8	0.3	Georgia	23	94.1
Indiana	27	96.6	0.3	Vermont	24	93.9
Wyoming	28	95.8	-2.5	Iowa	25	93.6
Vermont	29	94.9	-6.2	North Carolina	25	93.6
North Dakota	30	94.6	-2.8	Louisiana	26	93.4
Maine	31	94.0	2.4	Wyoming	26	93.4
South Dakota	32	92.9	-1.0	Tennessee	27	93.1
Montana	33	91.6	-5.3	New Mexico	28	92.9
Florida	34	90.6	-1.5	Utah	28	92.9
Georgia	35	90.0	-0.5	Missouri	28	92.9
Tennessee	36	89.9	2.5	Nebraska	28	92.9
North Carolina	37	89.6	1.4	Indiana	28	92.9
West Virginia	38	89.4	4.8	South Carolina	29	92.6
Kentucky	39	89.2	5.7	Oklahoma	30	92.5
Idaho	40	89.0	7.7	North Dakota	31	92.2
Arizona	41	88.0	11.3	Kentucky	32	92.1
Oklahoma	42	87.3	1.7	Maine	33	92.0
Texas	43	87.1	-0.4	Alabama	34	91.9
Alabama	44	86.9	0.3	West Virginia	34	91.9
Louisiana	45	86.8	-3.7	Kansas	35	91.7
South Carolina	46	84.9	-3.9	South Dakota	36	91.6
Arkansas	47	84.8	0.9	Idaho	37	91.3
Utah	48	84.8	14.2	Montana	38	91.1
New Mexico	49	83.6	12.1	Arkansas	39	90.9
Mississippi	50	81.6	4.8	Mississippi	40	90.3

[1]100 = national average for all states weighted by their population.
[2]Alaska and Hawaii are excluded from the study but assumed to rank first and second.

SOURCES: Walter W. McMahon, "Geographical Cost of Living Differences: An Update," MacArthur Spencer Series Number 7, Illinois State University, August 1988; and F. Howard Nelson, "An Interstate Cost-of-Living Index," in *Educational Evaluation and Policy Analysis*, Spring 1991, vol. 13 no. 1, pp. 103-111.

Poverty Thresholds Adjusted by State for Cost of Living: 1990

State or Region	One Person	One Person Under 65	One Person 65 and Over	Two Persons	Householder Under 65	Householder 65 and Older	Three Persons	Four Persons	Five Persons	Six Persons	Seven Persons	Eight Persons	Nine or More Persons
Alabama	5,625	5,750	5,300	7,195	7,436	6,684	8,610	11,296	13,353	15,084	17,115	19,095	22,702
Alaska**	7,700	7,871	7,255	9,849	10,178	9,149	12,060	15,463	18,279	20,649	23,429	26,139	31,076
Arizona	6,170	6,307	5,813	7,892	8,156	7,331	9,684	12,390	14,646	16,545	18,773	20,944	24,901
Arkansas	5,564	5,688	5,247	7,117	7,355	6,611	8,715	11,173	13,206	14,920	16,929	18,887	22,455
California	6,482	6,626	6,107	8,291	8,568	7,702	10,153	13,017	15,387	17,387	19,723	22,004	26,160
Colorado	6,109	6,244	5,755	7,813	8,075	7,258	9,568	12,267	14,501	16,381	18,587	20,736	24,654
Connecticut	7,700	7,871	7,255	9,849	10,178	9,149	12,060	15,463	18,279	20,649	23,429	26,139	31,076
Delaware	6,317	6,457	5,952	8,080	8,350	7,506	9,894	12,685	14,995	16,929	19,220	21,443	25,493
District of Columbia	7,408	7,571	6,978	9,473	9,790	8,800	11,600	14,873	17,581	19,861	22,535	25,141	29,891
Florida	5,956	6,086	5,611	7,618	7,872	7,077	9,328	11,960	14,138	15,971	18,121	20,217	24,036
Georgia	5,760	5,888	5,427	7,367	7,614	6,844	9,021	11,567	13,673	15,446	17,525	19,552	23,246
Hawaii**	7,700	7,871	7,255	9,849	10,178	9,149	12,060	15,463	18,279	20,649	23,429	26,139	31,076
Idaho	5,588	5,713	5,265	7,148	7,387	6,640	8,753	11,223	13,256	14,986	17,004	18,970	22,554
Illinois	5,962	6,094	5,617	7,625	7,881	7,084	9,338	11,972	14,152	15,967	18,140	20,236	24,061
Indiana	5,686	5,813	5,358	7,273	7,517	6,757	8,906	11,419	13,498	15,249	17,302	19,303	22,949
Iowa	5,729	5,857	5,398	7,328	7,573	6,808	8,973	11,505	13,600	15,364	17,422	19,448	23,122
Kansas	5,613	5,736	5,288	7,179	7,419	6,869	8,791	11,272	13,324	15,052	17,078	19,053	22,653
Kentucky	5,637	5,763	5,311	7,211	7,452	6,696	8,830	11,321	13,382	15,117	17,153	19,137	22,751
Louisiana	5,717	5,844	5,386	7,317	7,557	6,793	8,954	11,481	13,571	15,331	17,395	19,407	23,073
Maine	5,631	5,756	5,306	7,203	7,444	6,691	8,820	11,309	13,368	15,101	17,134	19,116	22,727
Maryland	6,782	6,933	6,390	8,675	8,965	8,058	10,622	13,620	16,099	18,187	20,635	23,022	27,371
Massachusetts	7,455	7,621	7,024	9,536	9,855	8,859	11,677	14,972	17,698	19,992	22,684	25,208	29,988
Michigan	5,839	5,969	5,502	7,469	7,719	6,926	9,146	11,727	13,862	15,659	17,767	19,822	23,567
Minnesota	5,852	5,982	5,513	7,485	7,735	6,953	9,165	11,751	13,891	15,699	17,805	19,864	23,616
Mississippi	5,527	5,650	5,208	7,070	7,306	6,568	8,657	11,100	13,121	14,822	16,817	18,763	22,307
Missouri	5,686	5,813	5,358	7,273	7,517	6,757	8,906	11,419	13,498	15,249	17,302	19,303	22,949
Montana	5,576	5,700	5,254	7,132	7,371	6,626	8,734	11,198	13,237	14,953	16,986	18,929	22,504
Nebraska	5,686	5,813	5,358	7,273	7,517	6,757	8,906	11,419	13,498	15,249	17,302	19,303	22,949
Nevada	6,005	6,138	5,657	7,680	7,937	7,135	9,405	12,058	14,254	16,102	18,270	20,383	24,234
New Hampshire	6,605	6,751	6,223	8,447	8,730	7,848	10,344	13,263	15,679	17,711	20,095	22,419	26,655
New Jersey	7,700	7,871	7,255	9,849	10,178	9,149	12,060	15,463	18,279	20,649	23,429	26,139	31,076
New Mexico	5,686	5,813	5,358	7,273	7,517	6,757	8,906	11,419	13,498	15,249	17,302	19,303	22,949
New York	6,929	7,083	6,528	8,862	9,159	8,233	10,852	13,915	16,448	18,581	21,082	23,521	27,964
North Carolina	5,729	5,857	5,398	7,328	7,573	6,808	8,973	11,505	13,600	15,364	17,432	19,448	23,122
North Dakota	5,844	5,969	5,517	7,318	7,460	6,706	8,839	11,333	13,397	15,134	17,171	19,157	22,776
Ohio	5,864	5,994	5,325	7,500	7,751	6,966	9,184	11,776	13,920	15,725	17,842	19,906	23,665
Oklahoma	5,642	5,786	5,334	7,242	7,484	6,726	8,868	11,370	13,440	15,163	17,227	19,220	22,850
Oregon	5,766	5,894	5,453	7,375	7,622	6,851	9,031	11,579	13,687	15,462	17,544	19,573	23,270
Pennsylvania	6,201	6,336	5,842	7,921	8,196	7,368	9,712	12,452	14,719	16,627	18,866	21,048	25,024
Rhode Island	6,537	6,682	6,159	8,361	8,641	7,788	10,339	13,128	15,518	17,530	19,890	22,191	26,383
South Carolina	5,668	5,794	5,340	7,250	7,492	6,735	8,878	11,382	13,458	15,199	17,246	19,240	22,875
South Dakota	5,607	5,731	5,283	7,171	7,411	6,662	8,782	11,259	13,309	15,035	17,060	19,033	22,628
Tennessee	5,699	5,825	5,369	7,289	7,533	6,771	8,925	11,444	13,527	15,281	17,329	19,344	22,998
Texas	5,760	5,888	5,427	7,367	7,614	6,844	9,021	11,567	13,673	15,446	17,525	19,552	23,246
Utah	5,686	5,813	5,358	7,273	7,517	6,757	8,906	11,419	13,498	15,249	17,302	19,303	22,949
Vermont	5,748	5,875	5,415	7,351	7,597	6,829	9,002	11,542	13,644	15,412	17,488	19,511	23,196
Virginia	5,943	6,076	5,600	7,602	7,856	7,062	9,309	11,936	14,109	15,936	18,084	20,175	23,987
Washington	5,937	6,069	5,594	7,594	7,848	7,055	9,299	11,923	14,094	15,922	18,065	20,155	23,962
West Virginia	5,625	5,750	5,300	7,195	7,436	6,684	8,810	11,296	13,353	15,084	17,115	19,095	22,702
Wisconsin	5,839	5,969	5,502	7,469	7,719	6,938	9,146	11,727	13,862	15,659	17,767	19,822	23,567
Wyoming	5,717	5,844	5,386	7,312	7,557	6,793	8,954	11,481	13,571	15,331	17,395	19,407	23,073

** Alaska and Hawaii are given the value for the highest state, Connecticut

CENTER FOR THE STUDY OF EDUCATIONAL FINANCE
Department of Educational Administration and Foundations
College of Education and the Graduate School
Illinois State University, Normal, Illinois

The MacArthur/Spencer Special Series on Illinois School Finance

MacArthur/Spencer Series Number 20

GEOGRAPHICAL COST OF LIVING DIFFERENCES: INTERSTATE AND INTRASTATE, UPDATE 1991

Walter W. McMahon
Shao-Chung Chang

With cartographic assistance from Kevin Carlock

Center for the Study of Educational Finance
Illinois State University
Normal, Illinois 61761

April 1991

Abstract

Geographical Cost of Living Differences:
Interstate and Intrastate, Update 1991

This paper develops a method for estimating current geographical differences in the cost of living index for all states for 1981-1990. These estimates based on BLS data are shown to correspond closely to statewide cost of living estimates for 1989 based on the American Chamber of Commerce Research Association data for selected cities.

The paper also develops estimates of the cost of living as among large cities, metropolitan areas, and nonmetropolitan areas within each state for 1989, and for all countries within Illinois for 1989.

Living costs are highest in Hawaii, Alaska, Connecticut, Washington D.C., New Jersey, Massachusetts, New York, and California. They are lower in Mississippi, West Virginia, Arkansas, Idaho, and Utah. There is a 57 percent difference in the purchasing power between the highest and lowest states, whereas the variation in real purchasing power within states between the higher cost large cities and lower cost metropolitan and nonmetropolitan areas is 22 to 35 percent respectively.

The basic pattern of differences persists since 1977 with shifts related to economic growth rates.

The authors are Professor of Economics, and graduate student in Economics respectively, at the University of Illinois at Urbana-Champaign. The latter was particularly involved in developing Part V on Intrastate differences. They would like to express appreciation to Dale Mitchell and David Braddock; to Wenhui Hu for able research assistance; as well as to Patric Hendershott, Donald Haurin, and to two anonymous referees for constructive suggestions. We also appreciate the assistance of Alan Hickrod and the MacArthur/Spencer Foundation who supported work on this project.

Geographical Cost of Living Differences: Interstate and Intrastate: Update 1991

Walter W. McMahon and
Shao Chung Chang

Significant differences in the cost of living exist among different parts of the country, as well as among different rural and urban counties within the same state. But no systematic estimates of differences in the cost of llving among states have been computed since 1980. The Bureau of Labor Statistics discontinued collecting and publishing its cost of living index for 24 SMSA's in 1981, and the American Chamber of Commerce estimates are also for selected cities and only for the most recent years.

A systematic procedure for estimating these differences among states and localities based on the Bureau of Labor Statistics data for the 24 SMSA's was developed earlier by McMahon and Melton (1978). The resulting estimates for 1977 found many uses. The basic method was adopted and extended by Fournier and Rasmussen (1986) who produced estimates by states for 1980. But since then there have been large differential impacts among states following the oil price increases of 1979-80, the recession and oil price declines in 1981-83, the effects of high interest rates on exchange rates and agricultural exporting states, and the industrial recovery in the late 1980s. All of these could be expected to lead to differential effects on prices and costs among geographical areas and therefore a changed pattern of geographical cost of living differences.

The ideal way to evaluate these differences would be to collect price data from PMSAs, MSAs, and Nonmetropolitan areas in every state, weighting these by the population in each area, and to also conduct detailed budget studies of family expenditures in each of these localities to establish the necessary geographical variation in the weights to be placed on each budget component. This procedure would be prohibitively expensive, however, and therefore likely will never be done in this detail. The Bureau of Labor Statistics in fact has moved in the opposite direction by discontinuing the collection and publication of its cost of living index in 1981. The American Chamber of Commerce Research Association ACCRA (1990) has recently started collecting data and computing an index for selected cities. But the budget weights do not vary by geographical area, and the index is not computed on a statewide or on a countrywide basis.

What is needed is a reduced form (predictive) equation that can be used to estimate the COL by states, or by counties within each state based on successive readings on the key explanatory variables in each place at each date, checking to see that the structure does not change. This paper does this, refining the procedure used in McMahon-Melton (1978) and in Fournier and Rasmussen (1986) to adapt it to both the kind of data that are available and to new housing value data that are now available on an annual basis. The results then are cross checked with statewide estimates based on the ACCRA (1990) sample. The result is an index of the

cost-of-living estimated for each state from 1981 through 1990 based on the predictor variables for each state both on a base year (U.S. average for 1981 = 100) and on a normalized annual (U.S. average for each year = 100) basis. The paper concludes with a brief consideration of the nature of changes in the geographical differences in the cost of living between the earlier studies for 1977 and 1980 and the present, as well as of the trends during the 1980's.

I. Existing Cost of Living Measures and Their Uses

Both the US BLS (1982) index for 1981 and the ACCRA (1990) index for 1989 are for selected cities, and the geographical boundaries of the relevant PMSAs and MSAs change over time. The Consumer Price Index (CPI) is published by the BLS for 15 major urban areas, as well as for urban and rural breakdowns within the four geographical regions for the U.S., but it has a base of 100 in the base period (1982-84 = 100) and therefore does not show the initial differences in the level of living costs. The CPI instead is an index of price changes since that time, whereas in fact the cost of living in the base year in these places varied considerably. The CPI also is not available by state.

The method adopted therefore seeks to take these base-year differences in the cost of living into account by starting with the last report for a family cost of living budget for 1981 as reported by the BLS (in 1982). Attention is confined to those explanatory variables that have a logical relationship to the cost of living within each of the MSAs, since as much stability as possible in their predictive capacity is sought, and also to variables for which data is available on an annual statewide basis for 1981-1990. For these reasons there are some differences in the explanatory variables from those used by Fournier and Rasmussen (1986, p. 184) who do not seek relationships to the structural parameters, or stability over time, because they are concerned only with the single 1980 census year. However, the explanatory variables used here do reflect the major factors used in the Fournier and Rasmussen (1986) analysis as well as in the original McMahon-Melton (1978) estimates for 1977. After exploring the relationships within these MSAs, the cost of living index is then generalized to a statewide basis as explained below, and the stability of the relationship to these same explanatory variables is explored. This procedure is better than using the MSAs in each year because as mentioned above the geographical boundaries of the MSAs change over time, causing problems for explanatory variables such as population change and per capita income, whereas the statewide measures of these variables can be expected to be considerably more accurate over time.

The rationale for uses of specific cost of living indexes is straightforward. Geographical differences in the cost of living affect the purchasing power of wages and salaries, which are always paid in nominal dollars, at different locations. For salaries to be comparable in real terms they therefore must be deflated (i.e., divided by) a geographical cost of living index such as the one developed here. To avoid questions of interpersonal comparisons of utility, the BLS concept

of a standard budget for a family of four, which we use here, is one that seeks to keep the head of the household on the same indifference curve with respect to commodities purchased irrespective of where he or she locates.

This concept, however, includes the living costs but does not include the non-monetary benefits of different locations (e.g., the sunshine, seaside, or access to alternative and better job opportunities), benefits that partly justify the higher costs and that also affect location decisions. That is, it may cost individuals more to maintain the same living standard in certain locations, but those locations may offer various additional advantages that they are willing to pay for. A geographical cost of living index is limited to differences in the monetary costs of living such as differences for comparable housing accommodations in different places, which can be substantial.

The uses that have developed for geographical cost of living indices, as well as an interpretation of its potential misuses, depend upon this concept. It is useful to employees in making decisions to locate because, to the extent that the cost side is to be considered in making these decisions, it is what the salary will buy in <u>real</u> terms, not in nominal terms, plus their evaluation of the non-monetary returns that basically govern the outcomes. That is, in analyzing the choice, the evidence is that a "money illusion" is not strong, after allowing for lags in adjustment, in which case employees would tend to make a correction for price level and cost of living differences first, and then evaluate the non-monetary benefits, albeit implicitly. Because of this behavior, multiplant firms with plants in different locations, state school systems with urban and rural unit districts, universities competing in interstate job markets, and other kinds of employers who wish to maintain salaries that are comparable in different locations (plus or minus the non-monetary environmental fringes) must also normally make some adjustment either explicitly or implicitly both for differences in costs of living as well as for the non-monetary advantages of the higher cost locations rather than looking only at the more purely nominal wage and salary differences. Some adjustments for non-monetary returns to particular locations or regions have been considered by Roback (1988) and by Blomquist et al. (1988).

A cost of living index has also been used to adjust production costs or investment costs to real terms when making geographical cost comparisons. This would include econometric estimates of cost functions using cross-section data, interstate comparisons of adequacy in education (e.g., A. Hickrod et al., 1987, p. 9), and comparisons of rates of return to education such as in the study by Israeli (1983) where the author extends the cost living index for the sampled population of 39 MSAs for an earlier year to the non-sampled population of 237 MSAs. A cost-of-living index is not precisely the same as the cost of production, investment costs, or an index of educational costs, but the procedure should give a reasonable approximation in those cases where wage and salary costs are a very large percentage of total costs, as is true in the case of schools and colleges for example. Geographical differences in the smaller

75-540 O - 94 - 3

non-labor costs in these cases may also be correlated with geographical differences in living costs, but this is a point that could be examined in special cases.

II. The Theory and The Model

There have been several earlier attempts to investigate the sources of differences in the cost of living in addition to the recent ones mentioned above. Sherwood (1975), for example, used the BLS indices and price data to construct standard budgets that isolate the effect of climatic differences on costs. But his indices are limited to this one source of differences and also were constructed for only the 44 cities and regions in his BLS sample. Haworth, Rasmussen, and Mattila (1973) and Alonso and Fajans (1970) explored the extent to which urban population and other variables explain differences in the cost of living within the BLS sample, but they did not undertake predictions for nonsampled areas. Alonso (1970) finds urban population size, when income is included, to be of minor significance. Israeli (1977) found that housing differences were a good predictor of the differential in nominal wages and prices among selected cities. But the only major efforts to extend cost of living indices from sampled to nonsampled areas have been by Simmons (1973, 1988) and by McMahon and Melton (1978). Simmons sampled prices in 12 Florida counties and then used regression equations to extend these prices to all counties in the state. The first result, in the absence of budget studies to obtain the necessary weights, is therefore closer to a geographic price index than to a cost of living index. Augmented by budget studies, it has been used by the State of Florida since 1974 in the Florida school aid formula. McMahon and Melton (1978) developed a model that explains cost of living differences within the BLS sample, and then used the regression coefficients, together with measures of the explanatory variables for the non-sampled areas, to extend the cost of living index to all 50 states and to estimate the cost of living for counties within California, Illinois, Pennsylvania, and Texas. Fournier and Rasmussen (1986) updated this as indicated above for states for 1980, but only for this one Census year. Now the data availability has changed, and there is need to update the index on an annual basis for the 1981-1990 period.

Economic theory suggests that changes in the effective demand for goods and for housing, especially when supplies are not perfectly elastic, can play a large part in the determination of geographical differences in living costs.

The demand function for goods and services in any given locality expresses the quantity demanded primarily as a negative function of price ($\alpha_1 < 0$), a positive function of per capita income in the locality ($\alpha_2 > 0$) and a positive function of both assets in housing, H, and imputed housing user costs ($\alpha_3 H$) which include capital gains and losses:

(1) $$q = a_1 p + a_2 Y + a_3 H + a_4 \Delta P + \mu_1$$

Here p = a price index relevant to goods and services purchased in the area,

q = a market basket of goods and services needed to sustain a family of four at the same level, irrespective of the area,

COL = pq = the cost of living,

Y = per capita personal income in the locality from U.S. Department of Commerce data,

H = value of the house of given size and quality (measured here as the median sales price of existing single-family homes available from the National Association of Realtors (1990),

ΔP = percent change in the population in the area over the preceding five years, and

μ_1 = disturbances.

The structural factors shifting the demand function, Y, H, and DP, have a logical basis in economic theory and can first be considered briefly. Individual income is a critical element in the demand for virtually all goods and services, raising demand by shifting the budget constraint outward when income is higher because most goods are normal goods ($a_2 > 0$). Where supply is inelastic (as in the case of land prices), especially for those items that are not transportable or geographically mobile, this can bid up the price and lead to geographical differences in living costs.

Consumer demand can also be increased by an asset effect, and the value of housing, H, is a significant component of total assets. The Life Cycle Hypothesis of Ando and Modigliani (1963), which with various extensions by Friedman, Heckman, and others dominates the theory of the household, measures it by using the total stock of assets or net worth. But such a comprehensive measure of all assets is less relevant for purposes of analysis of geographical price differences than are the assets specific to the locality in the form of equity in housing. Apart from this asset effect, it is also that land is immobile resulting in an inelastic supply, so that when demand rises, housing prices are driven up which means a higher imputed annual user-cost of housing. Sherwood (1975, p. 14) found that out-of-pocket housing costs vary widely among areas, ranging from an index of 168 in Boston to 68 in Austin, Texas. Using the median sales price of housing in a locality as an index to housing costs and as a measure of past asset accumulation that includes capital gains and losses has the further merit of being a measure that is widely available for all years for many large and small metropolitan areas on an annual basis from the National Association of Realtors (1990), whereas both housing costs and the more comprehensive asset measures are not.

Climatic differences also may have effects on differences in living costs. So we will explore below the merits of using an additional variable for climate, C.

Population growth has ambiguous effects on prices, as was stressed earlier by McMahon and Melton (1978, p. 326). Rapid population growth accompanied by effective purchasing power can increase the pressure on some facilities other than housing, and act to raise their prices (i.e., $\alpha_4 > 0$). However, per capita income is a better measure of effective demand, and because it and H are included as variables explaining per-unit costs, this effect of only population growth (that does not necessarily have the purchasing power) is less likely to be strong. On the other hand, economies of scale in certain services such as schools and city services also can be achieved as pointed out by Alonso (1970, pp. 72-75) (i.e., on the supply side below $\alpha_7 < 0$).[1] The net effect cannot be inferred from economic theory, but because of the large migration toward the south and the sun belt states during the 1980s where economies of scale could be meaningful, the hypothesis is that this relationship will be negative ($\alpha_4 < 0$).

The supply equation expresses price as a positive function of the quantity supplied both in the short run and in the long run ($\alpha_5 > 0$), as well as of housing costs as mentioned above ($\alpha_6 > 0$):

$$p = \alpha_5 Q + \alpha_6 H + \alpha_7 \Delta P + \mu_2 \tag{2}$$

where μ_2 = disturbances and all other variables have been defined under equation (1). Assuming linearity, the demand and supply functions may be solved simultaneously eliminating q. The resulting reduced-form price equation then can be multiplied throughout by the appropriate quantity weight $\bar{q}$ representing the market basket of commodities in the standard budget for a family of four. Because these quantity weights are designed to maintain the same level of well being in each area, they are treated as constants and as part of the parameters in equation (3) below. This result contains the key determinants of the cost of living, COL, in each locality:

$$COL = p\bar{q} = \frac{\alpha_2 \bar{q}}{1/\alpha_5 - \alpha_1} Y + \frac{(\alpha_3 + \alpha_6/\alpha_5)\bar{q}}{1/\alpha_5 - \alpha_1} H + \frac{(\alpha_4 + \alpha_7/\alpha_5)\bar{q}}{1/\alpha_5 - \alpha_1} \Delta P + \mu_3 \tag{3}$$

Because $\alpha_1 < 0$, all denominators can be expected to be positive. The first two numerators relating to Y and to H also can be expected to be positive as suggested above, and because the hypothesis is that $\alpha_7 > 0$, the sign of the third numerator is indeterminate.

III. Estimation of the Model

The parameters can be simplified as shown in equation (4), the model to be estimated. Here β_1 and β_2 are expected to be positive, and β_3 to be indeterminate, but probably negative since the positive effects of population increase on the demand side are likely to be picked up by Y and H, whereas the negative effects due to economies of scale and the movement further out and to retirement communities remain:

(4) $$COL = \beta_1 Y + \beta_2 H + \beta_3 \Delta P + \mu$$

The definitions and data sources for the variables are:

COL = Cost of Living Index, for 1981 for 24 MSAs and 4 regional non-metropolitan areas as published by the U.S. Bureau of Labor Statistics (1982, p. 45). These and the ACCRA (1990) measures for the respective sets of states within which sample data for selected localities exists are extended to a statewide basis in 1981 and 1989 respectively by using a weighted average of the MSA and non-metropolitan components of the COL. Weights consist of the percent of the population that is metropolitan vs. non-metropolitan in each state from the U.S. Bureau of the Census.

Y = Per Capita Personal Income, in thousands of dollars. For states this is from U.S. Department of Commerce Survey of Current Business (1990) where it is also available for these MSAs and by county within each state. (Disposable income is not available on an equally consistent basis.)

H = Value of Housing, measured as the median value of an existing one-family home. This is available from the Census of Housing, U.S. Department of Commerce for 1980 only, and from the National Association of Realtors (1990) as reported in the U.S. Statistical Abstract (Table 1236) for 1981-1990.

ΔP = Percent Change in Population, for the preceding five years, from Current Population Reports, Series P-25, U.S. Department of Commerce (1990, p. 16, Table 1), and various other issues.

The results obtained for the regression for the MSAs in 1981, the last year the BLS collected data, and for the regressions using statewide data for the corresponding states for 1981-1990 are shown in Table 1. The signs are all as expected, and the t-statistics indicate that all coefficients reach the 0.05 level of significance or above except for that on Y in a few of the earlier years where it is closer to the 0.10 level of significance. Multicollinearity among the explanatory variables is reasonably low (as shown in Appendix A), with the expected positive simple correlation between Y and H of .38 the highest. The R^2 as shown above is reasonably good for cross section data (and highest in the most recent years).

Turning to the statewide regressions (Eqns. 6-15), the procedure used is one of first constructing a statewide 1981 COL index for the states in which the MSAs are located by weighting the BLS index for the 24 (metropolitan) MSAs and their index for the nonmetropolitan areas by each states' distribution of population as between metropolitan and nonmetropolitan areas. In Table 1, a comparison of Eqs. (5) and (6) reveals regression coefficients for the MSA and statewide data that are very similar. In Appendix A, a test is shown to see if as between the two

75-540 O - 94 - 4

Table 1
Major Determinants of Cost of Living Differences
(t-statistics are in parentheses)

MSAs, n = 24:

		R^2
(5)	1981 COL = .0015Y + .178H - .57ΔP + 74.1 (0.952) (2.04) (-3.05)	.514

Statewide (Population Weighted) Means, Based on BLS Data, n=22:

(6)	1981 COL = .002Y + .182H - .56ΔP + 67.6 (1.63) (2.61) (-2.22)	.552
(7)	1982 COL = .002Y + .163H - .62ΔP + 74.4 (1.55) (1.87) (-2.17)	.463
(8)	1983 COL = .002Y + .191H - .65ΔP + 72.3 (1.89) (2.23) (-2.24)	.549
(9)	1984 COL = .002Y + .274H - .90ΔP + 72.4 (1.77) (3.19) (-2.74)	.635
(10)	1985 COL = .002Y + .285H - 1.12ΔP + 72.8 (2.34) (4.63) (-3.65)	.758
(11)	1986 COL = .002Y + .289H - 1.37ΔP + 74.2 (2.27) (5.40) (-4.21)	.811
(12)	1987 COL = .0014Y + .266H - 1.54ΔP + 83.9 (1.74) (4.96) (-3.70)	.806
(13)	1988 COL = .002Y + .202H - 1.62ΔP + 84.3 (2.54) (4.23) (-3.35)	.804
(14)	1989 COL = .002Y + .154H - 1.40ΔP + 75.3 (3.40) (3.37) (-2.62)	.778

Statewide (Population Weighted) Means, Based on ACCRA Data, n = 34

(15)	1989 COL = .002Y + .141H + .01ΔP + 62.5 (5.44) (7.48) (0.13)	.870

regressions there is a change in the structure. The null hypothesis cannot be rejected, indicating that there is no significant change. This is not surprising because the inference is that at least the first two variables are structural factors. Also a very high percentage of the population in almost all these 22 states is metropolitan as opposed to nonmetropolitan or rural.

When climate, C, is added to the regression measured as a dummy variable with a value of 0 below the Mason-Dixon line, including California, to reflect the lower heating costs in sun-belt states, especially while oil and related energy prices were very high, the result shown below in Eq. (16) is typical:

$$(16) \qquad 1981\ COL = \underset{(0.863)}{.0013Y} + \underset{(2.63)}{.233H} - \underset{(-1.64)}{.44\Delta P} + \underset{(1.09)}{5.21C} + 68.9 \qquad \frac{R^2}{.575}$$

Using "climate" would seem to have more logic than merely using regional dummies. But although the R^2 is slightly higher, the significance of each of the other explanatory variables is reduced in comparison to Eq. (6), and the t-statistic for climate never reaches the 0.05 level in any year from 1981 through 1990. So given this lack of significance, climate was dropped as an explanatory variable.

Other regressions were tested, using population <u>levels</u> in place of the change in the population over time, for example. But when per capita income is included as an explanatory variable, as shown in Eq. (16) below, population level, P, is not insignificant:

$$(17) \qquad 1981COL = \underset{(1.58)}{.002Y} + \underset{(2.39)}{.172H} - \underset{(-2.09)}{.55\Delta P} + \underset{(0.02)}{.00001P} + 67.1 \qquad R^2 = .564$$

The R^2 is no higher than when P is dropped, as in Eq. (6), and lower than when climate is used instead of P in Eq. (16). This insignificance of population levels when per capita income is included was discovered earlier by Alonso (1970).

It is impossible to test the regressions as shown for years prior to 1981 because the number of MSAs covered in the National Association of Realtors (and hence also the Statistical Abstracts) data on the median sales price of existing housing, H, diminishes and is totally inadequate. However, for further tests on the stability of the coefficients, the SMSA and statewide 1981 COL were updated for the years following 1981 (the latter shown in Table 1) by use of the Consumer Price Index, which shows percentage changes from the base year. For the 24 MSAs, the CPI is available, and the results of a second test for change in the structure from 1981 to 1989 is shown in Appendix A. This again reveals no evidence of significant structural change. For the statewide COL, the percentage increments from the base year in the CPI for metropolitan vs. nonmetropolitan areas were weighted by the percentage of the population living in metropolitan vs. nonmetropolitan areas in each state or region. This weighted percentage change in the CPI then was used to update the

base level 1981 COL. The CPI is based on budget studies that reflect the changing budget proportions over time in purchases in each area, and these CPI weights are periodically updated.[2] This method of updating in the sampled areas is also the method used earlier by the BLS to update their own index.

As a further check on the accuracy of the COL estimates in recent years, an independent data source for selected communities in 34 states sampled by the American Chamber of Commerce Research Association was used to create an ACCRA-based statewide COL index for 1989 for these states. The method is the same as that used to create the statewide index based on the BLS sample as described above. Specifically, all of the communities sampled by ACCRA within each state were grouped by PMSA, MSA, and nonmetropolitan areas (using neighboring states in those cases where there was no nonmetropolitan area sampled). The means within each category then were weighted by the proportion of the population in that state in PMSA's (if any), MSA's, and nonmetropolitan areas. The resulting weighted mean COL for each state in which sample data exists was used in a regression containing the same explanatory variables as shown in Eq. (15) in Table 1. There the significance of per capita income, Y, and housing, H, is very high, exceeding the .01 level. The regression coefficient for Y is the same and the coefficient for H is very similar to those in the BLS-based regressions. Population change, ΔP, however is not significant, with t = 0.13. Appendix A shows simple correlations among the explanatory variables that are in the same pattern (for this different sample) as for the BLS regressions. A test for differences in the coefficients does reveal a significant difference, undoubtedly due to the difference in the population change coefficient. The statewide COL estimates for the 50 states based on these independent BLS and ACCRA data sources will be compared and discussed shortly below.

All of the regression results suggest that differences in the median sales price of housing emerge as the most significant source of differences in the cost of living, although per capita income also is important. However, the median house prices, H, as reported by the National Association of Realtors overstates increases in constant-quality house prices by about 2 percent a year, as shown by Hendershott and Thibodeau (1990, pp. 328, 333). This overstatement is significantly related to changes in real income (see ibid., 1990). Therefore the increases in prices of a constant-quality house are likely to be somewhat less important, and larger per capita income somewhat more important as determinants of differences in the cost of living than the regression results might suggest, given that H is a fraction of Y. Direct out-of-pocket housing costs account for about 23 percent of a typical household budget, and the imputed own equity contribution due to capital gains and losses which vary together with H are likely to account for even more. Higher per capita income is also especially significant in Connecticut and the Northeast seaboard.

The effect of the percentage growth of population is not a major explanatory variable because its regression coefficient is multiplied by the very small values for ΔP as compared to Y and H, its effect is not

only smaller but also insignificant in the ACCRA-based regression. Nevertheless lower costs due to economies of scale in public services and perhaps movement by higher income and retired persons following tax cuts in the 1980s to places like New Hampshire (from Boston) and toward the retirement sunbelt states may still be a minor factor.

IV. Geographical Differences in the Cost of Living The Results

The differences in the cost of living among the 50 states and the District of Columbia are shown in Table 2 with the 1981 U.S. average treated as the base year. The index is obtained using the statewide regression equation (6) shown in Table 1 together with measures of per capita personal income and the median sales price of existing single family homes for each state and for each year from 1981-1990, as well as the percent change in population for the preceding five years for each state from 1977-1990.[3] The cost of living index then was normalized, with the results shown in Table 3, so that 100 represents the national unweighted average for each year for all states.[4]

These results in Table 3 indicate that there is a 42 percent variation in the cost of living in 1990 among states in the continental U.S., and a 57 percent variation if Hawaii and Alaska are included. The higher cost of living states continue to be in the East, Connecticut, New Jersey, and the District of Columbia in particular, plus Alaska and Hawaii. In these places higher incomes, higher prices, and higher housing costs are all a factor. The lower living cost states are those in the South, such as Mississippi, Arkansas, Alabama, South Carolina, Arizona, and New Mexico, where there are lower heating costs, and less population density may contribute to lower costs of land. The Midwestern and North Central states remain in the middle.

Table 4 shows the normalized cost of living index for 1989 based on the ACCRA regression (Eqn. 15) compared to the 1989 index based on the BLS data. These are rank ordered from highest to lowest cost using the ACCRA-based index. The percentage differences shown on the right are quite small, considering the differences in the concepts discussed below, with a difference of less than 3.3 percent between the estimates in 75 percent of the states. The differences range from 0 percent (when rounded) in North Carolina, Nebraska, Wyoming, Oregon, Texas and Utah to a high of -8.23 percent in Arizona, +7.56 percent in Missouri, +7.03 percent in Rhode Island, and -6.58 percent in New Mexico.

Examination of the reasons for these differences reveals three sources, that may be useful to those wishing to make evaluative judgments in the use of the results:

1. In a few states where there are very large cities, the ACCRA samples are sometimes confined to one suburb that may not be representative, e.g., Nassau-Suffolk to represent New York, Schaumburg for Chicago.

This appears most frequently to lead to a small over-estimate of costs by the ACCRA-based index.

2. In other states there are large cities that are not sampled by ACCRA (e.g., Providence, RI and Alexandra, VA), or that are grouped by the U.S. Census with the MSAs even though they contain more people (e.g., St. Louis and Kansas City, MO). Since they are underrepresented, this could lead to an underestimate of costs using the ACCRA-based index in these states.

3. Some states have had a huge influx of population in the five years leading up to 1989 (e.g., Arizona +16.7 percent, Florida + 14.6 percent, 7.1 percent in New Mexico, and 15.5 percent in New Hampshire). This could contribute to some understatement of the true cost of living by the BLS-based index for these places. There are no percentage losses of population in any state that are anywhere near this large.

4. Beyond this, there is the more general point that the ACCRA-based index uses the same budget weights on prices in all regions (e.g., no heavier weight for the higher heating costs in Maine). So the concept is slightly different, and the ACCRA index is perhaps somewhat closer to a geographical price index than to a cost of living index.

Therefore some differences in the BLS-based and ACCRA-based statewide indices are to be expected. But overall, the relatively small percentage differences in the two estimates, the very small differences in the range from highest to the lowest, and the reasonably close correspondence in the rank order serve to increase confidence in the accuracy of the estimates in Table 3, perhaps substituting the ACCRA-based statewide estimate in those three or four states indicated in point #3 above that have had extraordinarily large increases in population in the late 80's.

With respect to changes over time, the pattern remains much the same as in McMahon and Melton (1977). Living costs in Massachusetts, Connecticut, District of Columbia, Alaska and Hawaii, which were high in 1977, now are relatively even higher. And the lower cost of living areas such as Kentucky, Louisiana, Mississippi, West Virginia and Wyoming now are relatively even lower. The recession in the farm states throughout the 1980's lowered living costs there since 1977 relative to the other states, and the industrial recession in 1981-83 lowered per capita incomes and relative living costs in the industrial states. But then the later industrial recovery from 1983-89 also appears to have been a factor in raising demand and costs. In this recovery period, increases in the cost of living begin to occur in Massachusetts, Virginia, and parts of the industrial midwest that perhaps have been arrested by the 1991 recession.

Table 1
Geographical Cost of Living Index, by State, 1981-1990 *

COL	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
Connecticut	109.14	111.25	116.55	119.19	123.86	131.75	160.46	166.19	169.68	168.63
Maine	93.60	95.03	97.93	100.76	103.62	109.15	116.27	118.62	122.35	122.39
Massachusetts	106.76	108.75	108.66	116.37	120.18	127.06	133.28	138.69	142.22	143.77
New-Hampshire	96.02	96.33	99.50	105.09	108.45	113.99	119.13	126.13	125.67	129.25
Rhode-Island	99.03	100.64	103.32	106.26	108.63	111.56	121.21	129.13	127.57	139.29
Vermont	93.61	96.79	96.99	99.57	101.63	109.16	116.21	118.60	121.06	123.96
Delaware	102.56	104.02	106.83	109.65	112.51	116.10	133.05	128.76	138.26	130.50
Dis-Columbia	117.86	119.58	131.31	121.03	126.81	126.10	132.73	139.69	146.28	148.58
Maryland	102.61	105.22	107.36	110.06	113.01	116.96	118.26	121.86	126.06	128.28
New-Jersey	108.29	108.19	111.91	116.26	120.27	125.62	133.68	139.65	143.66	148.96
New-York	106.12	105.47	107.99	111.25	116.03	118.96	133.30	130.22	132.33	136.90
Pennsylvania	100.78	102.07	103.36	105.66	108.32	111.09	116.16	118.75	119.63	121.19
Illinois	102.85	104.30	105.97	108.90	110.24	112.67	116.09	118.36	121.60	126.08
Indiana	97.03	98.09	100.06	101.68	104.16	105.66	108.36	110.20	113.81	115.30
Michgan	98.62	99.77	102.09	105.09	107.76	109.98	111.37	113.35	116.16	118.61
Ohio	99.35	100.89	102.86	105.01	106.80	108.79	110.36	112.15	114.91	117.01
Wisconsin	98.59	100.11	101.77	104.37	106.10	108.12	109.73	111.17	116.01	117.93
Iowa	99.36	99.97	100.71	103.10	106.89	107.63	108.95	109.96	113.68	115.35
Kansas	97.24	100.62	101.67	101.11	105.72	108.15	110.06	111.93	119.89	115.27
Minnesota	101.73	103.10	104.87	108.09	110.19	112.61	116.52	116.10	118.96	120.05
Missouri	99.05	99.94	101.65	104.11	106.75	108.96	110.85	112.82	116.95	118.29
Nebraska	98.85	99.79	100.57	102.81	104.92	106.80	108.36	109.66	111.71	116.91
North-Dakota	96.77	97.16	97.59	99.90	100.16	103.71	106.08	105.70	109.61	110.62
South-Dakota	95.83	96.63	96.90	98.56	99.31	101.61	103.16	103.63	106.70	108.73
Alabama	92.28	93.79	95.78	97.93	99.00	101.19	102.07	106.50	107.38	108.60
Arkansas	91.30	93.09	94.96	97.37	99.57	100.85	101.50	102.96	105.65	107.25
Florida	91.19	92.97	96.62	99.03	102.04	104.33	106.63	108.73	113.23	116.78
Georgia	93.38	94.97	97.16	100.33	101.85	104.50	106.69	108.73	111.76	113.00
Kentucky	92.93	93.87	96.88	97.65	98.96	103.11	109.60	108.15	109.37	110.81
Louisiana	95.64	96.91	97.83	99.93	101.83	101.97	103.29	108.85	107.92	110.11
Mississippi	93.19	94.64	96.00	97.81	98.79	99.31	100.91	101.04	101.81	104.65
North-Carolina	95.95	97.56	99.67	102.12	104.14	106.97	110.19	111.65	115.37	117.36
South-Carolina	90.80	92.29	94.39	96.63	98.59	100.72	102.36	103.83	106.36	108.82
Tennessee	92.66	94.55	96.64	99.69	101.54	103.60	105.73	107.90	110.24	111.04
Virginia	99.07	100.97	104.63	107.78	111.27	116.16	122.98	127.13	130.01	131.35
West-Virginia	93.06	94.02	94.92	96.76	98.66	100.73	102.67	106.12	108.78	110.98
Arizona	91.76	92.99	96.80	98.37	100.19	101.79	103.65	106.82	106.65	108.34
New-Mexico	93.55	95.00	95.09	96.98	99.27	100.82	102.66	104.65	106.88	108.73
Oklahoma	95.06	95.09	95.05	97.98	100.46	102.02	105.15	108.38	110.26	112.66
Texas	94.25	95.67	95.97	99.17	101.74	102.59	104.99	108.56	111.29	111.70
Colorado	99.02	100.79	102.66	106.33	108.16	110.56	113.27	115.63	119.76	120.91
Idaho	88.76	89.92	92.06	94.98	96.39	98.61	101.18	103.67	105.80	108.67
Montana	94.30	95.21	96.43	97.55	97.56	100.38	102.13	105.38	108.60	110.95
Utah	88.91	89.21	89.96	93.00	95.96	98.21	100.59	102.75	105.59	106.69
Wyoming	89.51	89.25	90.89	95.55	100.07	102.57	106.65	111.38	112.51	115.05
California	108.43	109.72	111.56	114.20	117.66	120.90	125.03	133.79	140.16	143.92
Nevada	90.90	92.26	95.63	100.21	103.68	106.91	108.65	110.19	113.95	116.97
Oregon	93.12	95.44	99.20	102.69	104.90	106.52	107.96	109.57	111.72	115.37
Washington	95.13	97.01	100.15	103.76	106.12	108.67	109.07	111.52	115.59	130.37
Alaska	118.97	118.06	116.64	113.75	116.97	116.88	123.96	136.33	140.05	158.61
Hawaii	111.09	111.59	113.07	115.52	121.21	125.15	129.62	138.36	151.30	166.68
U.S. Avg (Unweighted)	97.79	99.14	100.97	103.71	106.14	108.96	112.16	115.25	118.66	120.94

* Base year is 1981, where the U.S. average of BLS SMSA index = 100.

Table 3

Normalized Geographical Cost of Living Index, by State, 1981-1990

(Unweighted mean for each year = 100)

Normalized Col	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
Connecticut	111.60	113.21	117.44	116.92	116.70	120.93	125.35	126.85	125.98	122.88
Maine	91.71	95.85	96.99	97.16	97.64	100.17	101.89	102.75	102.13	103.20
Massachusetts	107.11	107.67	107.59	110.28	113.23	116.61	118.85	120.24	119.87	118.05
New-Hampshire	96.16	97.17	98.54	102.10	102.18	106.52	106.23	107.70	105.93	103.57
Rhode-Island	101.25	101.52	103.73	103.64	102.15	104.23	108.08	108.58	107.52	106.91
Vermont	95.72	95.62	96.05	96.00	97.66	100.37	101.85	102.91	102.72	101.32
Delaware	96.98	104.92	105.80	105.73	106.02	106.55	108.37	109.12	108.11	107.91
Dis-Columbia	120.53	120.61	120.14	118.62	117.59	115.73	118.25	123.11	127.27	122.86
Maryland	104.92	104.14	104.13	104.13	104.48	105.69	105.65	105.57	104.25	104.07
New-Jersey	108.68	109.17	110.81	112.08	113.32	115.11	119.30	121.17	121.07	120.89
New-York	105.67	105.39	105.95	107.37	108.19	109.16	110.02	111.26	111.53	111.56
Pennsyvania	101.06	102.96	102.36	101.86	102.06	101.96	101.79	101.30	100.75	100.19
Illinois	105.17	105.20	104.95	104.62	103.86	101.61	102.45	102.70	102.69	102.60
Indiana	99.21	98.94	99.10	98.81	98.16	97.71	96.67	95.69	95.67	95.41
Michgan	100.86	100.63	101.11	101.32	101.50	100.96	99.31	98.36	97.89	97.81
Ohio	101.59	101.76	101.84	101.25	100.62	99.85	98.39	97.31	96.56	96.75
Wisconsin	100.81	100.98	100.79	100.56	100.06	99.23	97.84	98.66	97.78	97.51
Iowa	101.60	100.86	99.76	99.48	98.82	98.79	97.15	95.39	96.99	95.30
Kansas	99.44	101.50	100.69	99.63	99.61	99.36	98.12	97.12	98.83	96.32
Minnesota	104.07	104.20	103.85	104.32	103.81	103.17	103.12	100.98	100.27	99.93
Missouri	100.27	100.81	100.67	100.60	100.87	100.01	98.66	97.89	98.88	96.16
Nebraska	101.08	100.66	99.60	99.13	98.85	99.02	96.52	95.13	96.16	95.03
North-Dakota	98.96	97.98	96.85	95.37	94.36	94.26	91.80	91.72	93.22	91.67
South-Dakota	98.00	97.37	95.97	95.01	93.57	93.35	91.97	89.83	89.96	89.91
Alabama	94.38	94.60	94.85	94.43	93.35	93.07	91.73	90.67	90.61	89.80
Arkansas	93.16	93.90	94.05	93.79	93.81	93.55	90.61	89.22	88.88	88.69
Florida	95.39	93.77	95.69	95.68	96.11	95.65	96.90	96.36	96.58	96.91
Georgia	95.69	95.80	96.20	96.66	95.96	95.98	95.13	96.33	96.20	92.61
Kentucky	95.03	94.68	93.96	94.16	93.22	94.67	94.16	93.84	93.18	91.63
Louisiana	97.81	97.75	96.98	96.37	95.97	93.59	93.19	91.86	90.97	91.08
Mississippi	95.70	95.46	95.08	94.31	93.07	91.16	89.62	88.56	87.51	86.52
North-Carolina	98.12	98.60	98.71	98.46	98.12	98.18	98.22	96.87	97.35	97.02
South-Coroline	93.85	93.09	93.68	93.18	92.89	92.44	91.25	90.09	89.65	89.81
Tennessee	94.55	95.17	95.70	96.13	95.67	94.90	94.27	93.62	93.91	91.91
Virginia	101.31	101.85	101.62	102.92	106.83	108.61	109.66	110.21	109.58	109.61
West-Virginia	94.12	94.85	94.00	93.28	93.92	93.63	91.55	91.21	91.69	91.76
Arizona	93.83	93.80	93.89	94.85	94.60	93.62	92.36	90.95	89.89	89.80
New-Mexico	95.66	95.83	94.18	93.51	93.52	93.53	91.52	90.87	90.09	89.93
Oklehoma	97.21	96.72	96.16	96.67	96.65	93.86	93.76	93.95	93.91	93.99
Texas	96.38	96.30	95.05	95.63	95.86	96.13	93.62	96.19	93.91	94.01
Colorado	101.26	101.66	101.67	102.63	101.90	101.67	101.01	100.15	100.08	99.98
Idsho	90.74	90.70	91.86	91.58	90.81	90.12	90.22	89.78	89.18	89.85
Montana	96.47	96.04	95.50	96.06	93.91	93.12	91.95	91.64	91.54	91.76
Utah	90.92	88.98	88.89	89.67	90.61	90.15	89.66	89.16	89.93	90.21
Wyoming	90.51	90.03	90.01	92.13	96.38	96.16	95.10	96.55	96.87	95.16
Californie	110.86	110.68	110.66	110.11	110.86	110.98	111.68	110.22	110.16	110.01
Nevada	92.95	93.06	94.50	96.83	97.68	98.13	96.88	96.61	95.20	96.72
Oregon	95.32	96.17	98.25	99.03	98.74	97.77	96.25	95.07	96.16	95.60
Washington	97.37	97.85	99.39	100.01	99.98	99.56	97.97	96.76	97.67	98.81
Alaske	121.66	119.08	115.52	109.68	108.32	107.27	110.56	110.29	125.66	121.18
Hawaii	114.62	112.56	112.77	111.29	114.20	114.86	115.59	120.04	127.66	126.17
U S Avg (Unweighted)	100	100	100	100	100	100	100	100	100	100

Table 4
ida COL Index via Prediction Equation and ACCRA Sample Compared
1989

P.E. ank	STATE	Prediction Equation Normalized COL	ACCRA Normalized COL	ACCRA COL Non-normalized	Difference between P.E. and ACCRA Normalized COL
3	Alaska	125.46	129.15	133.14	-2.85%
1	Hawaii	127.44	128.52	132.50	-0.84%
2	Connecticut	125.98	125.22	129.09	0.61%
4	Dis-Columbia	123.27	121.71	125.48	1.28%
5	New-Jersey	121.07	118.39	122.05	2.26%
6	Massachusetts	119.87	115.73	119.31	3.58%
8	New-York	111.53	115.44	119.01	-3.39%
7	California	118.14	113.66	117.18	3.94%
13	New-Hampshire	105.93	110.04	113.44	-3.74%
10	Delaware	108.11	109.25	112.63	-1.05%
17	Pennsyvania	100.75	107.88	111.22	-6.62%
12	Maryland	106.25	104.60	107.84	1.57%
9	Virginia	109.58	103.86	107.05	5.53%
16	Illinois	102.49	103.27	106.46	-0.75%
20	Michigan	97.89	103.05	106.24	-5.01%
31	Florida	96.59	100.68	103.79	-6.04%
11	Rhode-Island	107.52	100.46	103.57	7.03%
22	Washington	97.43	99.01	102.07	-1.59%
14	Maine	103.13	98.13	101.17	5.09%
46	Arizona	89.89	97.95	100.98	-8.23%
15	Vermont	102.72	97.44	100.45	5.42%
23	North-Carolina	97.25	97.34	100.35	-0.10%
28	Nevada	95.20	97.04	100.04	-1.89%
44	New-Mexico	90.09	96.44	99.42	-6.58%
21	Wisconsin	97.78	95.76	98.72	2.12%
18	Minnesota	100.27	95.47	98.42	5.03%
32	Georgia	94.20	95.19	98.13	-1.03%
19	Colorado	100.08	95.13	98.07	5.21%
25	Ohio	96.86	94.90	97.84	2.06%
34	Nebraska	94.16	94.59	97.52	-0.46%
30	Wyoming	94.83	94.39	97.31	0.47%
33	Oregon	94.16	94.29	97.21	-0.14%
42	Louisiana	90.97	94.23	97.15	-3.47%
35	Texas	93.81	93.86	96.76	-0.05%
27	Indiana	95.67	92.85	96.75	1.94%
38	North-Dakota	92.22	93.65	96.55	-1.53%
41	Montana	91.54	93.60	96.50	-2.21%
29	Iowa	94.98	93.30	96.19	1.79%
43	Alabama	90.51	92.79	95.66	-2.46%
45	South-Dakota	89.94	92.65	95.52	-2.93%
47	South-Carolina	89.65	92.09	94.94	-2.65%
26	Kansas	95.83	91.79	94.63	4.40%
36	Tennessee	92.91	91.68	94.52	1.34%
49	Arkansas	88.88	91.54	94.37	-2.91%
48	Idaho	89.18	90.62	93.42	-1.59%
39	Kentucky	92.18	90.60	93.40	1.75%
51	Mississippi	87.51	90.50	93.30	-3.30%
37	Oklahoma	92.91	90.21	93.00	3.00%
24	Missouri	96.89	90.07	92.86	7.56%
40	West-Virginia	91.69	89.95	92.72	1.94%
50	Utah	98.83	99.06	91.81	-0.25%
U.S.	Unweighted Ave.	100.00	100.00	103.09	0.022

V. Cost Differences Within States

Walter W. McMahon and Shao Chung Chang

Differences in the cost of living among large cities, medium sized cities, and nonmetropolitan areas in each state are developed in Table 5 based on the ACCRA data collected for selected cities. These results will be compared to cost of living estimates for each county in Illinois (Table 6) and metropolitan nonmetropolitan differences based on these independent BLS-based regressions (in Table 7). The results of the two independent approaches again are reinforcing. But the results also again reveal some locations where there appear to be sampling errors in the means that are based on the ACCRA data.

Metropolitan-Nonmetropolitan Differences

To estimate differences in the cost of living between larger cities and nonmetropolitan areas, all of the locations sampled by ACCRA were first grouped within each state by PMSA (> 1.5 million population), MSA (1.5m-50,000) and nonmetropolitan areas (< 50,000). Since many states have no cities large enough to be a PMSA, there are blanks for these states in column 1 of Table 5. Where there were no nonmetropolitan area data collected (see * in Column 3), or where the ACCRA data is for only one location (see c in Table 5) the state was pooled with data for the same city size class in adjacent states. For some of the very largest cities such as San Francisco and Chicago, ACCRA has collected data for only one suburb (marked b in column 1, Table 5). In these cases a regression equation based on housing values, per capita personal income, and population change is used to obtain the estimate shown. Estimates at this level of detail should be used with some caution, and with one eye to the cities in which data was actually collected by ACCRA because there is some variation within each size category. Nevertheless certain patterns emerge.

The cost of living is distinctly higher in larger cities, and only slightly higher in the medium sized metropolitan areas from that in nonmetropolitan towns of < 50,000. There is a 74 percent difference in the cost of living as between the higher cost cities and the lowest nonmetropolitan areas in the U.S. This is larger than the 57 percent variation in the state averages in Table 3, as might be expected. But even within the same state, the real purchasing power of the persons living in the largest cities is 22 to 35 percent below the purchasing power of those living in medium sized cities and nonmetropolitan areas respectively.

Table 5 Cost of Living Index, 1989
For Large Cities, Metropolitan Areas, and Non-Metropolitan Areas

STATE	Large City (Pop > 1.5 m)	MSA's (1.5 m.-50,000)	Non-metropolitan (Pop < 50,000)
Alabama		96.02	94.90 a
Alaska		127.40	137.10
Arizona		101.13	100.43
Arkansas		96.30	93.10
California		110.79	99.29 a
Anaheim-Santa Ana	130.90		
Los Angeles-Long Beach	129.20		
Riverside-San Bernardino	110.24		
San Francisco	151.84 b		
San Jose	129.90		
Colorado		99.63	93.45
Denver	102.10		
Connecticut		131.75	99.33 a
Delaware		112.85 c	102.80
Dis-Columbia		125.50 c	
Florida		101.58	97.20
Miami-Hialeah	113.50		
Georgia		98.95 c	98.30
Hawaii		132.50 a	132.50
Idaho		96.10	92.75
Illinois		109.64	97.35
Chicago	120.10 b		
Indiana		96.77	93.46 c
Iowa		96.30	93.95
Kansas		98.85	89.80
Kentucky		95.97	91.20
Louisiana		98.80	93.45 a
Maine		104.00 a	99.30 a
Maryland		108.30	101.80 a
Massachusetts		120.29	99.30 a
Michigan		104.93	103.50
Detroit	117.63 b		
Minnesota		100.03	93.23 a
Mississippi		96.02 a	93.30
Missouri		94.45	88.95
Montana		93.41 a	93.84 a
Nebraska		92.45	89.33
Nevada		104.87 d	104.40 d
New-Hampshire		122.30	99.33 a
New-Jersey		122.03 c	122.03 c
Newark	122.03 c		
New-Mexico		100.83	98.04
New-York		109.82	99.90 c
Nassau-Suffolk	137.73 b		
New York	131.48 b		
North-Carolina		99.19	96.80
North-Dakota		98.60	93.23
Ohio		98.29	96.07
Cleveland	111.94 a		
Oklahoma		93.75	87.00
Oregon		99.00	94.90
Pennsylvania		104.60	99.50
Philadelphia	129.20		
Pittsburgh	106.10		
Rhode-Island		103.94 a	99.23 a
South-Carolina		96.40	92.70
South-Dakota		98.90	94.95
Tennessee		95.30	92.93
Texas		99.89	98.09
Dallas	104.20		
Houston	99.10		
Utah		92.10	90.80
Vermont		103.96 a	99.33 a
Virginia		113.27	101.80
Washington		97.42	92.70
Seattle	113.20		
West-Virginia		93.87	92.07 a
Wisconsin		99.80	98.10
Wyoming		93.81 c	93.84 c

*. Data is not available, so the index uses data from an adjacent state (or city).

a. Data is the same as Alabama, because there are no MSA's in Mississippi.

b. COL predicted using regression equation based on BLS sample, as explained in McMahon (1991). It uses data on housing values, per capita personal income, and population change specific to each large city. The resulting prediction for each city indicated (b) is before normalization to a statewide base of 100. To accomplish this adjustment, a regression equation was computed in each case for a neighboring city that does have ACCRA data, and the ratio of the BLS based prediction to the ACCRA estimate in the neighboring city is used to "normalize" the BLS-equation predictions to the same base.

c. The data presented by ACCRA data is incomplete and is not representative, or is missing, so the regional index for the respective MSA's or Non-metropolitan areas is used.

d. For Nevada MSA's and Non-metropolitan areas respectively, 1989 and 1990 ACCRA data is pooled.

By Counties

Differences in the cost of living by counties in Illinois have been estimated using the BLS-based regression equation and data on 1989 per capita income, housing values, and population change for each county. The results shown in Table 6 are normalized to a statewide average of 100 using first a mean weighted by the population of each county, and then an unweighted mean giving equal weight to each county. The population weighted mean is more relevant where expenditure are being distributed (as in a state school aid formula), whereas the unweighted mean would be more relevant to an individual trying to decide whether or not to move from one location to another. However, when the counties are rank ordered, the rank order is totally unaffected by the type of mean that is used for normalization, and the percentage difference from the highest to the lowest is not significantly affected.

The pattern of estimated cost of living differences within Illinois is illustrated in the map in Figure 1 (normalized using the unweighted mean). There is a 62 percent variation from the highest (Lake County) to the lowest (Johnson County) cost of living location, close to the 57 percent variation among the state-wide averages but smaller than the 74 percent variation between the larger cities (San Francisco, New York) and the lowest cost nonmetropolitan areas (in Missouri, Kansas, and Oklahoma).

When Illinois counties are grouped by PMSA, MSA, and nonmetropolitan areas as shown in Table 7, and rank ordered by cost of living within each group, the pattern discussed above based on the ACCRA data again emerges. The population-weighted mean within each group of counties is within 1.6 percent of the ACCRA-based estimates in all cases as shown at the bottom of Table 7.

VI. Potential Applications to Education in Illinois

The implication of using regional cost differences such as those presented here based in the cost of living in state school aid formulas requires brief comment.

The cost of living index could be made specific to each school district using the regression equation presented in this paper based on the 1989 ACCRA data (Eqn 15). Personal income per capita is available by school district based on state income tax returns. For 1990 the taxpayer was given a list of four digit school codes which greatly improves the accuracy of the reporting. Percentage change in population is available only by county, but since this variable is not significant, the county-wide change may be a suitable proxy. Median house prices are not available by school district for recent years, but the Housing Census for 1990-91 may make these available eventually. The alternatives to this would be to use the county cost of living estimates (Table 6) or the large city metropolitan-nonmetropolitan averages relevant to each school district.

Table 6

Cost of living by County in Illinois

COUNTY	1989 Normalized Weighted COL	1989 Normalized Unweighted COL	COUNTY	1989 Normalized Weighted COL	1989 Normalized Unweighted
Adams	90.73	102.53	Lee	89.27	100.89
Alexander	78.23	88.41	Livingston	90.19	101.92
Bond	84.82	95.86	Logan	87.35	98.71
Boone	93.94	106.17	Macon	92.48	104.51
Brown	89.07	100.66	Macoupin	85.05	96.11
Bureau	90.86	102.68	Madison	89.43	101.06
Calhoun	84.09	95.04	Marion	84.91	95.96
Carroll	88.49	100.00	Marshall	92.88	104.97
Cass	86.16	97.37	Mason	88.19	99.67
Champaign	90.39	102.15	Massac	80.05	90.46
Christian	87.44	98.82	McDonough	85.68	96.83
Clark	83.46	94.32	McHenry	95.84	108.31
Clay	84.32	95.29	McLean	92.19	104.19
Clinton	89.28	100.90	Menard	88.92	100.49
Coles	85.37	96.48	Mercer	89.13	100.73
Cook	105.32	119.03	Monroe	98.60	111.43
Crawford	89.04	100.63	Montogomery	85.08	96.15
Cumberland	81.98	92.64	Morgan	88.55	100.07
De Kalb	92.92	105.02	Moultrie	86.93	98.24
De Witt	87.54	98.93	Ogle	92.52	104.56
Douglas	84.38	95.36	Peoria	95.88	108.35
Du Page	113.54	128.32	Perry	85.91	97.09
Edgar	86.07	97.27	Piatt	91.01	102.85
Edwards	85.38	96.49	Pike	83.47	94.33
Effingham	86.99	98.31	Pope	76.90	86.90
Fayette	82.24	92.94	Pulaski	76.18	86.09
Ford	91.26	103.14	Putnam	94.87	107.21
Franklin	83.73	94.62	Randolph	83.70	94.59
Fulton	88.19	99.66	Richland	88.80	100.35
Gallatin	84.23	95.19	Rock Island	93.60	105.78
Greene	80.99	91.53	Saline	87.01	98.34
Grundy	95.52	107.95	Sangamon	93.86	106.07
Hamilton	81.41	92.01	Schuyler	85.19	96.27
Hancock	86.10	97.31	Scott	84.41	95.39
Hardin	78.06	88.22	Shelby	84.51	95.52
Henderson	83.63	94.51	Stark	92.37	104.39
Henry	90.76	102.57	Stephenson	93.61	105.79
Iroquios	87.56	98.95	St. Clair	86.97	98.29
Jackson	85.49	96.62	Tazewell	93.92	106.14
Jasper	85.34	96.45	Union	83.24	94.07
Jefferson	87.62	99.02	Vermilion	86.69	97.97
Jersey	84.22	95.18	Wabash	88.26	99.75
Jo Daviess	88.80	100.35	Warren	89.05	100.64
Johnson	70.42	79.58	Washington	87.44	98.82
Kane	100.05	113.07	Wayne	87.33	98.70
Kankakee	89.16	100.76	White	87.20	98.54
Kendall	102.30	115.61	Whiteside	90.22	101.96
Knox	84.42	95.41	Will	95.90	108.38
La Salle	89.88	101.58	Williamson	84.12	95.06
Lake	114.39	129.27	Winnebago	95.28	107.68
Lawrence	87.87	99.30	Woodford	92.00	103.97

Population-Weighted COL Mean 100

Unweighted COL Mean 100

Table 7

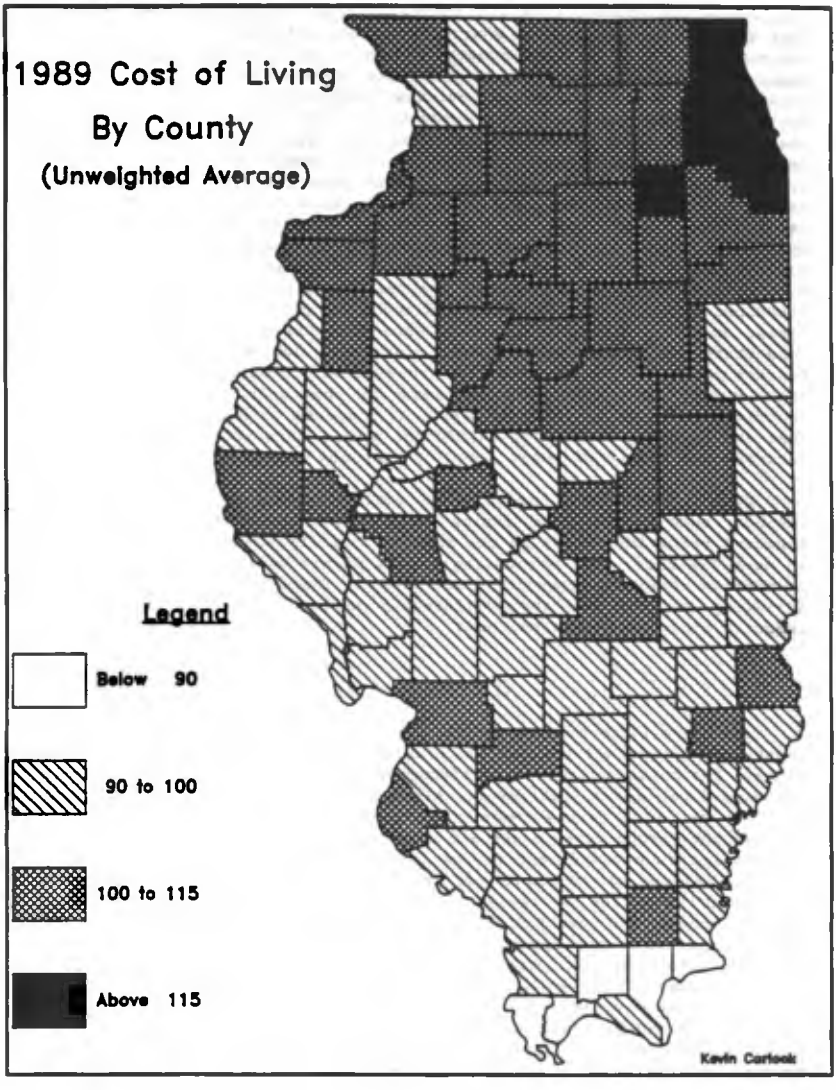
Regional Cost Differences Within Illinois
Comparison of BLS-Based County and ACCRA-Based Estimates
for PMSAs, MSAs, and Nonmet. Areas

As shown on the Illinois map (Figure 1)

Rank	COUNTY	1989 Normalized Unweighted COL	1988 Pop	% of Total Pop		Rank	COUNTY	1989 Normalized/ Unweighted COL	1988 Pop	% of Total Pop	
	PMSA										
1	Lake	129.27	493300	6.69	859.88	44	Carroll	100.00	17500	0.88	87.93
2	Du Page	128.32	760800	10.32	1311.09	47	Wabash	99.79	12800	0.89	69.16
3	Cook	119.03	5284300	78.97	9444.52	48	Mason	99.67	17400	0.87	87.12
4	Kendall	113.81	38800	0.52	59.92	49	Fulton	99.66	38500	1.93	192.78
5	Kane	113.07	316900	4.25	481.06	50	Lawrence	99.36	16900	0.88	84.22
7	Will	108.28	348700	4.68	506.63	51	Jefferson	99.02	37400	1.88	186.87
9	McHenry	106.21	171100	2.30	248.89	52	Iroquois	98.95	31600	1.59	157.09
10	Grundy	107.95	32300	0.44	47.12	54	Christian	98.82	35600	1.79	176.75
	Total Population		7446100	100.00		55	Washington	98.52	15600	0.78	77.45
	Population-Weighted Mean				119.39	56	Logan	98.71	31600	1.59	156.72
						57	Wayne	98.70	17900	0.89	88.26
	MSA					58	White	98.54	17300	0.86	85.15
6	Monroe	111.63	23200	1.02	112.87	59	Saline	98.36	27900	1.40	137.54
8	Peoria	106.35	182700	8.39	900.83	60	Effingham	98.31	32800	1.64	161.01
11	Winnebago	107.48	252100	11.57	1244.22	62	Moultrie	98.26	14400	0.72	71.07
13	Boone	106.17	30100	1.28	146.71	63	Vermilion	97.97	90600	4.35	445.96
14	Tazewell	106.14	124800	5.73	608.11	64	Cass	97.37	13900	0.70	68.00
15	Sangamon	106.07	180000	8.26	876.54	65	Hancock	97.31	22900	1.15	111.44
16	Stephenson	105.79	49300	2.26	239.43	66	Edgar	97.27	20560	1.03	100.18
17	Rock Island	105.78	155400	7.14	753.46	67	Perry	97.09	21700	1.09	105.55
21	Ogle	104.81	122700	5.68	593.53	68	McDonough	96.83	34200	1.72	166.38
23	DeKalb	104.19	116700	5.72	596.46	69	Jackson	96.62	59600	2.99	289.31
26	Woodford	102.97	22900	1.31	157.03	70	Edwards	96.49	7900	0.40	38.90
28	Henry	102.57	53300	2.65	316.98	71	Coles	96.48	52300	2.62	252.02
30	Champaign	102.13	172100	7.98	807.08	72	Jasper	96.69	11700	0.86	54.27
34	Madison	101.08	252900	11.58	1170.63	73	Schuyler	96.27	7000	0.29	27.73
35	Clinton	100.90	36200	1.57	158.62	74	Montgomery	96.15	31900	1.68	164.11
37	Kankakee	100.74	97300	4.49	452.85	75	Macoupin	96.11	49200	2.47	237.58
41	St. Clair	98.29	269700	12.36	1216.97	76	Marion	95.98	43200	2.18	208.79
44	Jersey	95.19	20500	0.95	90.01	77	Bond	95.85	16100	0.81	77.84
	Total Population		2175200	100.00		78	Shelby	95.82	22600	1.19	113.23
	Population-Weighted Mean				103.89	79	Knox	95.61	54200	2.82	269.60
						80	Scott	95.39	6100	0.31	29.26
	NONMETROPOLITAN AREAS					81	Douglas	95.26	19600	0.99	93.90
12	Putnam	107.31	5700	0.29	30.70	82	Clay	95.29	14300	0.73	71.36
18	De Kalb	101.02	76000	3.82	400.99	83	Gallatin	95.19	7200	0.36	34.46
19	Marshall	104.97	13200	0.66	69.61	85	Williamson	95.04	58200	2.92	277.97
20	Ogle	104.54	44200	2.22	232.70	86	Calhoun	95.04	5600	0.28	26.74
22	Stark	104.39	6600	0.33	34.62	87	Franklin	94.83	42000	2.11	199.66
25	Ford	103.24	14700	0.74	76.17	88	Randolph	94.39	29500	1.79	169.18
28	Platt	102.83	16200	0.81	83.71	89	Henderson	94.31	8900	0.43	41.26
27	Bureau	102.68	36800	1.85	189.85	90	Pike	94.33	17900	0.90	84.86
29	Adams	102.53	67400	3.40	348.23	91	Clark	94.32	16600	0.82	77.72
31	Whiteside	101.95	62500	3.16	320.16	92	Union	94.07	18000	0.90	85.08
32	Livingston	101.92	40600	2.02	206.88	93	Fayette	92.94	21900	1.10	102.26
33	La Salle	101.58	107300	5.39	547.59	94	Cumberland	92.44	10900	0.53	50.72
36	Lee	100.89	34800	1.79	176.39	95	Hamilton	92.01	8500	0.45	41.16
38	Mercer	100.73	19200	0.91	92.10	96	Greene	91.53	16000	0.80	73.58
39	Brown	100.66	5000	0.25	25.29	97	Massac	90.46	15000	0.79	68.17
40	Warren	100.64	20100	1.01	101.62	98	Alexander	88.41	11300	0.58	51.08
41	Crawford	100.61	20100	1.01	101.62	99	Hardin	88.22	5300	0.27	23.69
42	Menard	100.49	11700	0.59	59.07	100	Pope	86.90	4300	0.23	18.77
43	Jo Daviess	100.35	23100	1.16	116.47	101	Pulaski	86.09	8900	0.43	36.77
46	Richland	100.35	16900	0.85	85.20	102	Johnson	79.98	11300	0.57	45.18
45	Morgan	100.07	37200	1.87	187.53		Total Population		1990400	100.00	
				BLS Eqn.			Population-Weighted Mean				98.37

	Non-Normalized	County-Based (above)	ACCRA-Based (Table 9)
PMSA	119.39	119.39	120.13*
MSA	103.89	103.89	109.96
NONMETROPOLITAN AREA	98.37	98.37	97.35

*Base regression equation as explained in footnote b, Table 5. The ACCRA data reports only Schaumburg (COL=126), which is not likely to be representative.


1989 Cost of Living
By County
(Unweighted Average)
Legend
Below 90
90 to 100
100 to 115
Above 115
Kevin Cartoski

There are however pros and cons of making adjustments for regional cost differences, especially where the current differences in expenditure per pupil (and loss of pupil equity) are as large as they are in Illinois. Regional cost differences reflect the costs to teachers and administrators of living in each area, as well as other geographical price differences. One way in which such an index would most likely be used, is to convert the "nominal" expenditure per pupil in each district to "real" terms, removing the main effects of regional price differences (that can be done by dividing expenditure per pupil in each district by the index). For example, the $13,600 per pupil spent in Winnetka, when converted to real terms (constant dollars), is approximately $11,889, whereas the $2,500 spent per pupil in a "poor" district, since costs are lower there, converts to about $2,874 in "real terms." If an adjustment for regional cost differences were introduced into the school aid formula, in the ways described above, to adjust all expenditures per pupil to a constant dollar basis, and nothing else were done, then the effect would be to provide more state aid to the higher income districts, and less state aid to the poorest districts. Since there is a considerable problem in Illinois with pupil equity, as dramatically portrayed by the range from the wealthier districts spending $13,600 per pupil, as compared to the poorer districts that are spending $2,000, or $2,500 per pupil, this act alone would just increase the amount of pupil inequity that now exists.

Other compensatory adjustments could be made in the school-aid formula, such as using per capita personal income rather than equalized assessed valuations (EAV) as the means of ability-to-pay of the families in each district, and introducing a much higher state-financed floor or foundation level of expenditure per pupil. Then the adverse effects of adjusting for regional cost differences on the current level of pupil inequity would be counterbalanced.

VII. Conclusions

There are large differences of 57 percent in the cost of living among states and 35 percent between large urban and smaller cities within each state. The basic pattern of differences between higher costs in Eastern Seaboard urban areas, California, Alaska, and Hawaii, and lower costs in Southern and rural areas tends to persist over time. This is largely because the larger urban areas and bedroom suburbs are typified by higher residential land costs, and higher fuel and other housing costs. These are also related to the higher per capita incomes. There may also be some nonmonetary benefits of living in these areas that at least partially justify some of the cost differences. But over time recent changes in the geographical patterns appear to be related to the 1983-89 industrial recovery affecting the northeast, which will likely be moderated by the 1990-91 recession. Lower oil prices later in the '80's affected the south in a different way, and the continuing farm recession holds living costs lower in the midwest farm states. In 1980-85 the industrial states were hurt more severely than the oil producing and western states. But prices appear to have been somewhat inflexible downward there, and these areas also re-covered more quickly than the agricultural states and rural areas, where land and housing prices remain somewhat lower.

Part of the income differences among areas--roughly a third--are purely nominal differences in monetary salaries, given that there are differences in the cost of living. In the absence of a money illusion, employers as well as employees interested in maintaining a parity between services that are purchased or provided in different areas within states or between states must make some kind of adjustment implicitly for differences in the cost of living as well as in nonmonetary amenities. A geographical cost of living index is one step toward making such adjustments somewhat more explicit.

NOTES

[1]Population change and the price of housing may both be endogenous to a limited extent. That is, with respect to population change, persons may be attracted to areas where living costs are lower (e.g., plants locating in Tennessee or the rural south). But the data on migration suggest that this effect is small in relation to the movements towards the sunbelt and to outside the suburbs by more affluent and retired people during the 80s (e.g., to Florida, Arizona, New Mexico, Nevada, and New Hampshire). The ACCRA regressions in Table 4 however suggest that even this effect is insignificant by 1989. With respect to the median price of housing, this is to some extent a function of per capita income (see Hendershott and Thibodeau, 1990). But to treat population change and median house prices as endogenous would require specification of a number of additional factors affecting population change other than the COL, and affecting housing demand other than just Y, going considerably beyond the scope of this paper.

[2]There is no alternative to this BLS method for updating cost of living comparisons over time. These data are routinely used in studies of geographic cost of living variations.

[3]A xerox of the rather large data set that underlies these predictions is available from the author on written request enclosing $5 to cover the cost of processing and mailing. The data for per capita personal income and population change are available for all states. The values of H are the mean of the large city and smaller metropolitan areas that are sampled within each state, maintaining consistency insofar as possible throughout the 1981-90 period (since more areas appear in the data in the later years). For the few states where there are no values of H in the National Realtors Association (1990) data, values from the 1980 Housing Census were used to establish a ratio of housing values to those in an adjacent state for which there are good data. Assuming this ratio to adjacent areas remains unchanged, the values for the missing locations were then estimated.

[4]The results shown use the simple unweighted mean, which would be the index relevant for individuals considering moving. An alternative normalization was done using a mean computed by weighting the COL index for each state by its population. Governments or firms allocating funds in ways that depend on the total population of each area (e.g., school districts) are likely to find the population-weighted mean used for normalization more relevant. The rank order of the COL among states however is unaffected, and the percentage difference in the COL among areas is not significantly affected by use of the population weighted vs. nonweighted mean. Table 3a containing the normalized COL by setting the population weighted mean equal to 100 is available from the author on request.

References

Alonso, William, and M. Fajars (1970), "Cost of Living by Urban Size," Institute of Urban and Regional Development, University of California, Working Paper 128 (Berkeley), 19 pp.

Alonso, William (1970), "The Economics of Urban Size," Regional Science Association Papers, XXVI, pp. 67-83.

American Chamber of Commerce Researchers Assn (1990). ACCRA Cost of Living Index, Fourth Quarter, 1989, Vol. 22, No. 4, Louisville Chamber of Commerce, Louisville, KY.

Ando, Albert, and Franco Modigliani (1963), "The 'Life Cycle' Hypothesis of Saving: Aggregate Implications and Tests," American Economic Review, Vol. LIII (March), as reprinted in AEA Readings in Business Cycles, pp. 398-426.

Blomquist, Glenn, Mark Berger, and John Hoehn (1988), "New Estimates of the Quality of Life in Urban Areas," American Economic Review (March), pp. 88-107.

Barro, Stephen M. (1981), "Educational Price Indices: A State of the Art Review," AIU Policy Research, 1701 K Street, N.W., Washington, D.C. (March), pp. 1-41.

Fournier, Gary M., and David W. Rasmusen (1986), "Salaries in Public Education: The Impact of Geographical Cost of Living Differentials," Public Finance Quarterly, Vol. 14, No. 2 (April 1986), pp. 179-198.

Haworth, C. T., and D. W. Rasmussen (1973), "Determinants of Metropolitan Cost of Living Variations," Southern Economic Journal, XL (October), pp. 182-92; J. P. Mattila (1976), "Comment," SEJ, XLII (April); Haworth & Rasmussen (1976), "Reply," ibid.

Handershott, Patric H. and T. G. Thibodeau (1990), "The Relationships Between Median and Constant Quality House Prices: Implications for Setting FHA Loan Limits," AREUEA Journal, Vol. 18, No. 3, pp. 323-333.

Hickrod, G. Alan et al. (1987), "Documenting a Disaster: Equity and Adequacy in Illinois School Finance, 1973 through 1988," MacArthur/Spencer Series No. 4, Center for the Study of Educational Finance, Illinois State University, Normal, Illinois (December), 43 pp.

Israeli, O. (1983), "Effect of Variations in Cost of Living and City Size on the Rate of Return to Schooling," Quarterly Review of Economics and Business, Vol. 23, No. 4, Winter 1983.

Israeli, O. (1977), "Differentials in Nominal Wages and Prices Between Cities," Urban Studies, Oct. 1977, 14, pp. 175-90.

McMahon, Walter W., and Carroll Melton (1978), "Measuring Cost of Living Variation," Industrial Relations, Vol. 17, No. 3 (October), pp. 324-332.

McMahon, W. W. and Carroll Melton (1977), "A Cost of Living Index for Illinois Counties and School Districts," in Perspectives on Illinois School Finance, Carol E. Hanes, ed., Illinois Office of Education, Springfield (November 1977), pp. 74-113.

National Association of Realtors® (1990), "Existing Home Sales," Realtors Journal, March 1990, July 1987, and earlier issues, Washington, D.C.

Roback, Jennifer (1988), "Wages, Rates, and Amenities: Differences Among Workers and Regions," Economic Inquiry (January), pp. 23-42.

Simmons, James et al. (1973), Florida Cost of Living Research Study, Tallahassee, Fla.: State University of Florida.

Sherwood, Mark (1975), "Family Budgets and Geographic Differences in Price Levels," Monthly Labor Review, XCVIII (April), pp. 8-15.

U.S. Bureau of Labor Statistics (1982), "Family Budgets for 1981: Final Report," Monthly Labor Review, (also all later issues for the CPI), U.S. Department of Labor, Washington, D.C.

U.S. Department of Commerce (1990), "Projections of the Population of States by Age, Sex, and Race, 1989-2010," Current Population Reports, Series P-25, No. 1035 and earlier issues, Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

U.S. Department of Commerce (1980), Census of Housing, Bureau of the Census, Washington, D.C.

U.S. Department of Commerce (1990), Survey of Current Business, and various earlier issues, Washington, D.C.

U.S. Department of Commerce (1988), "Current Population Reports," Bureau of the Census, Series P-25, No. 1010, Washington, D.C.

Ward, James G. (1987), "The Concept of Adequacy in Illinois School Finance," MacArthur/Spencer Series No. 5, Center for the Study of Educational Finance, Illinois State University, Normal, Illinois, 21 pp.

Ward, James G. (1988), "City Schools, Rural Schools," MacArthur/Spencer Series No. 6, Center for the Study of Educational Finance, Illinois State University, Normal, Illinois, 8 pp.

Appendix A
Simple Correlations Among the Explanatory Variables

Correlation Matrix: 1981 Statewide BLS-Based Regression:

		Y	ΔP	H	C	P
Per Capita Income	Y	1.00	-.24	.38	.31	.22
5-Yr. Population Change	ΔP		1.00	.22	-.53	.08
Value of Housing	H			1.00	-.47	-.00
Climate (1=Northern)	C				1.00	-.27
Population Level						1.00

Correlation Matrix: 1989 Statewide ACCRA-Based Regression: (n=34)

Y	1.00	.01	.58	.43	.32
ΔP		1.00	.02	-.31	-.07
H			1.00	-.13	.24
C				1.00	-.11
P					1.00

Tests for Change in the Structure

1981 and 1989 Statewide BLS-Based Regressions:

$$F = \frac{(ESS_R - ESS_{UR})/k}{ESS_{UR}/(N+M-2k)} = \frac{(2100-1811)/4}{1811/(22+22-8)} = 1.436$$

$$F < F_{4,34} \text{ (95\%)}$$

Therefore we cannot reject the null hypothesis (i.e., the coefficients are not significantly different; there is no evidence of change in the structure).

1981 MSAs and 1981 Statewide Regressions:

$F = 0.230$ $\quad F_{4,38} = 2.626$ (95% level)

$F < F_{4,38}$, null hypothesis not rejected (i.e., no significant difference in the structure.

1981 BLS-Based and 1989 ACCRA-Based Statewide Regressions

$F = 11.32$, the null hypothesis must be rejected. At least one of the coefficients (undoubtedly the coefficient for ΔP) between the two equations is significantly different.

Appendix B

The 1981 cost of living for MSAs updated by use of the changes in the CPI from the base year lead to the following regression results. The problem in using only MSA data is that the geographical boundaries for the MSAs change over time, and people who work in the MSAs settle outside the MSA boundaries. The result is that the change in population (ΔP) for the MSA is misleading, as is the per capita income (Y), and the regressions, although in the same pattern as those in Table 1, are less meaningful.

		R^2
(B-1)	1982 COL = .002Y + .090H - .72ΔP (1.18) (.89) (-3.37)	.417
(B-2)	1983 COL = .002Y + .131H - .82ΔP (1.33) (1.09) (-3.38)	.505
(B-3)	1984 COL = .002Y + .23H - 1.22ΔP (1.40) (1.77) (-3.08)	.540

Another problem is that the per capita income (Y) for the Honolulu MSA is misleading. At any given time, there are many relatively well to do American and Japanese tourists occupying expensive beach front hotels, as well as villas in Oahu outside Honolulu. The per capita income of these persons and the value of the hotels undoubtedly affect the local cost of living, even though their income is not included in the measures of per capita income for the Honolulu MSA. To at least partially avoid the distortion that this causes, the per capita income for San Francisco, which is slightly above the per capita income of full-time Honolulu residents, was used in all of the BLS-based regressions to reflect the higher per capita income of tourists from the continental U.S. and of persons living outside the SMSA who work in Honolulu and affect the cost of living there. The alternative of deleting Hawaii from the sample seems less desirable since all the rest of the data for Hawaii is reasonable and deletion would further reduce the sample size, raising the standard errors. As shown in Table 4, the estimate for Hawaii using this BLS-based prediction equation is very close (-0.84%) to the ACCRA-based estimate.

U.S. DEPARTMENT OF EDUCATION
OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT

NATIONAL CENTER FOR EDUCATION STATISTICS

SEP 28 1998

Honorable Dale E. Kildee
Chairman, Committee
on Education and Labor
House of Representatives
Washington, DC 20515-6101

Dear Mr. Chairman:

I would like to thank you and Chairman Sawyer for the opportunity to testify before your joint committee hearing. I apologize for my tardiness.in responding to your questions. I have provided answers for most of the questions addressed to me. There were several questions I did not address because the Census Bureau would be the most direct source of answers to these questions.

If I can be of any further assistance, please do oot hesitate to contact me.

Sincerely,

Emerson J. Elliott
Commissioner

Enclosure

1. **"Additional questions for Emerson Elliott"**

Q. What steps is the Department of Education taking to obtain school district boundary information from the counties in California that have neglected to participate so far in the census mapping project.

A. At present, we are taking no action to obtain school district boundary information for the non-participating California counties. It must be remembered that the California State Department of Education declined to participate in the School District Mapping project. Only with the intervention of Dr. Anne Hafner, then an employee of the South West Regional Educational Research Laboratory, were the counties afforded the opportunity to map their districts and provide the boundaries to the Bureau of the Census. Some chose not to avail themselves of this opportunity. Only through action by the California State Department of Education or by the non-participating counties themselves could the missing boundary information be provided.

"Sample questions for Census Bureau and NCES"

"Data from 1990 Census for LEA'S"

1.
Q. What is your judgment regarding the appropriateness of using the 1990 Census LEA data as the primary basis for making Chapter 1 basic or concentration grants. Are the LEA data sufficiently reliable for this purpose?

A. Unless something is done to improve the reliability of the estimates for the thousands of small districts, the LEA data are not sufficiently reliable to be used as the primary basis for making Chapter 1 basic or concentration grants.

2.
Q. What special problems have you encountered in attempting to compile the 1990 Census for California LEAs? Are the California data of the same quality as that for other states except for a few non-participating counties? How might LEA data now be obtained for any counties that did not earlier prepare LEA maps for use by the Census Bureau and NCES?

A. Other than the problems we had in getting the mapping done, we have had no special problems in compiling the data for California. The data appear to be of comparable quality as similar size districts in other States. On the other hand, in the other 49 States, the state education agencies provided preliminary data to the school districts which then had an opportunity to question and challenge those data. The State Education Agency in California did not circulate these preliminary data and there were no challenges from California.

3.
Q. If you feel the LEA data are not sufficiently reliable to be used by the federal government as the primary basis for making Chapter 1 basic and concentration grants, what is your basis for this determination? Why are the LEA data less reliable than county data for making Chapter 1 allocations.

What has been gained from the expenditure of an estimated $6 million of this compilation if the resulting data are insufficiently reliable to serve as the primary basis for Chapter 1?

A. I believe that the data are not sufficiently reliable because the amount of random (sampling) error is too large for a large number of districts. Counties, on average, are much larger (in terms of population) than school districts. Therefore, the number of counties with large sampling error on the estimate of poor children is small in comparison with school districts.

Although there is a substantial amount of random error in many of the district estimates of children in poverty, some 40 or so states do make use of them in various ways in making the subcounty allocations for Chapter 1 grants. In addition to their use for fund allocatinns, the large special tabulation of the census by school district is the basis for the Congressionally mandated report "concerning the social and economic status of children who reside in the areas served by different educational agencies". There is also keen interest in this tabulation by States, LEA's and educational researchers for addressing numerous issues.

4.

Q. Specifically with respect to sub-county allocations of Chapter 1 basic and concentration grants, which are now made by the states using a variety of indicators of low income, what would be the advantages and disadvantages of using 1990 census LEA data on poor children for sub-county allocations compared to the current use of data such as free and reduced school lunch participation?

A. The advantage nf using 1990 Census data for LEA's to do the subcounty allocations is that you would get rid of various biases which can creep into the process if you use nther data. For example, if the rate of application for the lunch programs for similar families varies across districts, then the use of the lunch program numbers will under-allocate money to a district where poor families tend not to apply for some reason. The cnst nf getting rid of this bias is the introduction of more random error. If all or the vast majority of districts in a State are fairly large, the amount of random error introduced will be small and, therefnre, there is a real advantage to using the census LEA data rather than the lunch program data. If there are many small districts, using the school lunch data may result in a more equitable distribution among the various districts. At present, each State is free to make this choice given the particular situation in their State.

5.

Q. If the 1990 Census LEA data were sufficiently reliable, would it not be preferable to use them rather than county data to allocate Chapter 1 basic and especially concentration grants, so that aid can be more precisely targeted on high poverty areas?

With respect to concentration grants in particular, would it not be desirable to direct grants to the small-to-medium size, high poverty LEAs that do not now receive grants because they are in low poverty counties, or to central city LEAs that now are combined with their surrounding, more affluent, suburbs when treated on a county basis.

A. Whether or not the allocation of basic grants would be improved depends upon the amount of error that is reduced by eliminating the bias compared to the amount of random error which is introduced. If there were minimal random error, then it would generally be preferable to use the census estimates. With respect to concentration grants, the random error associated with sampling will affect the targeting of small to medium size LEA's if the census estimates are used.

6.

Q. Are there ways in which the 1990 Census LEA data could be used as a partial basis for improving the target of Chapter 1 grants on high poverty areas without switching completely to making allocations by LEA from the Federal level? For example, might grants be calculated at an LEA level, but states be given discretion over the actual allocation of funds among the smallest LEAs?

A. We do not see any practical way to use the 1990 Census LEA data on a partial basis to improve targeting of funds on high poverty areas. However, we believe that the Department's reauthorization proposal would address the situation posed in your question. Under our proposal, the Department would continue to determine county allocations based primarily on census data. A State would have the option to distribute Basic and Concentration Grant funds directly to LEAs without regard to counties using the best available measure of poverty it chooses. The State educational agency (SEA) would have the flexibility to use a more current poverty measure such as counts from the Aid for Families with Dependent Children (AFDC) program or the free/reduced-price lunch program to allocate the funds it receives under the Federal formula directly to LEAs. The State would not have to rely on district level census estimates that quickly have become out-of-date and present questions of reliability for small school districts because of random variability inherent in the one-in-six sample used by the Bureau of the Census. The current law requires that an SEA allocating funds directly to LEAs use the same factors as the Federal formula, which locks it into using census mapping data. Our proposal would only require that an SEA using this procedure demonstrate to the Secretary that doing so will result in a reasonable allocation of funds.

7.

Q. Criticism of the potential accuracy of the data compiled from the 1990 Census for LEAs has focused on the small population size of many LEAs and resulting likelihood of population sampling errors, since only one out of six households is surveyed overall with respect to income. While the average size of LEAs is inevitably smaller than that of counties, since there approximately 5 times as many LEAs as counties. In 1990, 772 counties--25% of all counties--had fewer than 2,000 school-age children, while 309 counties--almost 10%--had fewer than 1,000 such children. What are the 95% confidence intervals for your pool school-age child estimates for some of these small counties? Are the Census data for these counties also "unreliable"?

A. (to be answered by the Census Bureau)

8.

Q. In your testimony, you mentioned the James-Stein method as a potential way to reduce the sampling error for small jurisdictions. Would this be effective in rural areas? Would it be helpful to combine small population areas in some cases? And how would you choose a threshold in size in which the numbers are adequately reliable?

A. One of the nice properties about the James-Stein (J-S) procedures is that one can develop them in all areas. They would be most effective for LEA's which have a large amount of sampling error. Thus the estimates for small districts, be they urban or rural, would be the most improved. The James-Stein estimate is a weighted average of two other estimates. One is the census sample estimates. The other is an estimate based on the overall relationship between the census sample estimate and an administrative number which is thought to be a good predictor of what the census is trying to measure. In this case, children receiving free or reduced price school lunch might be thought to be a good predictor of "school age children in poverty families." For large districts, the error in the census sample estimate is small. In such cases, its weight will be large and the J-S estimate will be very close to the census estimate. On the other hand, The J-S estimate for very small districts will be dominated by the other estimate because the census estimate is unreliable and therefore it will have a small weight. Thus there is no need to choose a threshold or make a judgment of which sample estimates are "adequately reliable."

"Census Data Update Techniques"
"Questions for any witness"

1.
Q. What would be the most reliable and cost-effective way to update Census data on children in poor families, whether at the county level or the LEA level, between now and the next decennial Census, so that we can minimize the likelihood of large population shifts, and attendant Chapter 1 allocation changes in the future?

A. (should be answered by the Census Bureau)

2.
Q. There is legislation that would require changing county data on poor school age children by the statewide rate of increase or decrease in all school-age children. What is your opinion of this proposal? Is its implicit assumption--that local trends in school-age poor children closely track statewide changes in total school-age population--valid? Are there states where, during the 1980s, total and poor school age populations moved in opposite directions?

A. In general, using a State rate of change as a proxy for the rate of change in a given county is a procedure which is used as a last resort. It is employed as a special case or as a fall-back procedure because of a special circumstance. If the Census Bureau is not able to develop acceptable county level updates, Congress may want to consider a formula which allocates Chapter 1 money to States rather than county areas. States would then be responsible for the allocation to individual school districts based on Congressional Department of Education guidelines and the best available and local data. The Census Bureau should be able to produce annually, acceptable estimates in the change in children in poverty families for States. Most, if not all, States will then work out a solution for updating the sub-State allocations to reflect changing economic conditinns.

3.
Q. What would be the advantages and disadvantages of conducting a mid-decade census in 1995?

A. (should be answered by the Census Bureau)

For Witnesses from Census Bureau and NCES

1. What is the current status of the proposal that has been worked on by the Census Bureau staff to provide updated county-level estimates of children in poor families biannually beginning in 1995? Does the Department of Education intend to support and contribute to this project? What are current estimates of its cost, and the probable reliability of the population estimate it would produce?

   The Department of Education is very interested in this project to update the decennial estimates. We have pledged to provide $100,000 per year for several years to help get the project moving. The remaining points of the question should be answered by the Census Bureau.

2.
Q. Do you keep up-to-date the LEA maps prepared for the purpose of compiling the 1990 Census by LEA?

A. The LEA maps that served as a basis for the purpose of compiling the 1990 Census by LEA were paper maps provided by the Bureau of the Census on which the state education agencies entered the school district boundaries. Those boundaries have been converted to "TIGER" Line files, a computer-based system of digitized geography. The computer tapes containing these digitized boundaries are being provided to each state. Each state may up-date the TIGER Line files to reflect changes in school district boundaries. NCES plans to set up a program later in the decade to have each State update its boundaries for a special tabulation of the 2000 Census.

Question Asked by Congressman Gunderson during the Hearing:

1.

Q. To your knowledge, is there any state that doesn't have some kind of system for determining on a regular basis the income levels of people by school district?

A. Based on discussion with knowledgeable people in several states as well as the Council of Chief State School Officers, I found no evidence that states routinely have systems for determining, on a regular basis, the income level of people by school district. On the basis of these discussions, I think that:

1) it is unlikely that many state education agencies collect income data themselves;

2) within some states, education uses district-level income data collected by another agency in allocating state aid;

3) other states may have the capacity to collect income data through tax rolls or to provide "poverty" estimates from AFDC rolls.

○