$2 / 5 / 6$
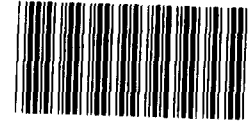
118121

INSTITUTE FOR PROGRAM
EVALUATION

B-206178

APRIL 19, 1982

The Honorable John Heinz
Chairman, Special Committee on Aging
United States Senate

118121

Dear Mr. Chairman:

  Subject: Identifying Computer Data Bases to Aid
       Congressional Oversight on the Needs
       of Older Americans (IPE-82-6)

  On April 30, 1981, you asked the Committee on Rules and
Administration (Senate Computer Center) and the General Account-
ing Office to assist the Senate Special Committee on Aging in
developing a national data base to aid oversight on the needs of
older Americans. The Senate Computer Center was to provide data-
processing support and GAO was to provide technical assistance in
the selection of existing data files. The fine support and coop-
eration of Mr. Larry Atkins, your staff economist, were invaluable
in this endeavor.

  In the July 15, 1981, Coordinating Conference, it was agreed
that GAO would provide technical assistance for two design tasks.
The first design task identified data archives and repositories
that emphasized "data on aging." It was completed on September 9,
1981, by the transmittal of a Working Paper entitled "Accessing
Data Archives and Repositories: Emphasis on Automated Data Per-
taining to Older Americans." The second involved selection of
data files that contain information of specific interest to the
Special Committee. Our enclosure completes this second and final
task of the assignment. Here, we summarize the scope and method-
ology of our overall effort.

  Because data files pertaining to issues on aging are not
housed in one central location, we set out to inventory sources
for the Special Committee. We applied two major criteria for
screening data archives and repositories. First, the data had
to have the potential for providing information on the four
policy areas of interest to the Special Committee--economics
(including retirement income and patterns of employment), health,
housing (including weatherization and fuel assistance), and demo-
graphics (including population trends). Our second criterion was
that the archive or repository had to be established and currently
accessible.

021642

(973165)

Using these two criteria, we identified twelve archives or repositories of potential use to the Special Committee. For each, our Working Paper included a summary of its activities, an overview of its machine-readable holdings, the name of a contact person, the cost for the computer tapes or services, turnaround time, and bibliographic references.

The second design task was to identify national data files of interest to the Special Committee. The original focus was on geographically based data files. Pursuant to discussions with your office, we agreed to identify a set of data files based on the individual or household as well.

For a first subset of national data files, the county was the most suitable geographic level because of its compatibility with the Senate Computer Center's Program Review System. In our review of the county-level data files at the Senate Computer Center, we found that only two sources provided information for the Center's geographic file--the County and City Data Book and fiscal data from the Community Services Administration.

To identify additional county-level data files, we

--obtained for the Special Committee the Area Resource File from the Health Resources Administration (the health data in this computer file are geographically based) and

--assessed the suitability of each county-level data file listed in GAO's Federal Information Sources and Systems (PAD-80-50, 1980), basing our subjective assessment on the abstracts contained in that directory.

We transmitted our findings in a document entitled "Senate Special Committee on Aging: County-Level Data Files," which was appended to our October 16, 1981, letter to Mr. Atkins.

Compiling a second subset of national data files based on the individual or household was more complex. At the outset, consideration was given to expanding the four broad policy areas (economics, health, housing, and demographics) to seventy specific issues. Our preliminary investigation revealed that cross-classifying numerous data files on each of seventy issues would be very time-consuming, contrary to the Special Committee's need for rapid turnaround. In our November 3, 1981, letter to Mr. Atkins, we suggested that the Special Committee consider thirteen important technical points in narrowing the range of automated data files to examine. Each of the thirteen points was designed to ask a question about the underlying structure of a data file and to highlight the essential information that should be known before its cross-classification on specific policy issues. Other technical issues might be raised, but we perceived these thirteen as being the most important for the Special Committee's purpose.

In the November letter, we also recommended an alternative approach to identifying data files based on the individual--namely, consulting experts in the four broad policy areas and asking their opinions on the most important data files. Accordingly, on November 24, 1981, we provided a list of more than fifty Data Specialists, from which your staff designated twenty for us to contact--five in each of the four broad policy areas.

In the enclosure to this letter, we summarize the results of our telephone survey of the twenty data specialists. We also present a separate list of data files recommended for each of the four areas and note additional considerations that the experts raised. We believe that this information will enhance the development of the Special Committee's national data base.

We hope that the products we have described and presented will help you increase your access to existing data files on the Nation's elderly. We appreciate the opportunity to have worked with you by providing technical assistance on this important task.

Sincerely yours,

Eleanor Chelimsky
Director

Enclosure

DATA FILES BASED ON THE INDIVIDUAL
OR THE HOUSEHOLD

The Senate Special Committee on Aging requested a list of data files based on the individual or the household. We used two approaches in developing this list. In the first, we considered categorizing the large number of relevant data files on each of the seventy policy issues. Experience told us that identifying a subset of automated files this way would be extremely time-consuming and might not be as useful as had been initially assumed.

In the second approach, we asked five data specialists in each of four broad policy areas--economics, health, housing, and demographics--to recommend existing data files that might be useful for policy analysis. For each policy area, we compiled a separate list of appropriate data files. These lists are summarized below, and we also note additional considerations offered by the five experts.

FIRST APPROACH--FORMING A MATRIX

Our first approach sought to develop a matrix or chart that would cross-classify each policy area with all relevant data files. At the outset, consideration was given to expanding the four broad policy areas (economics, health, housing, and demographics) to seventy specific issues. Working with a subset of files to produce the matrix, we began to have serious reservations. The large number of potentially useful data files meant that categorizing all relevant files would be very time-consuming, contrary to the Special Committee's expressed request for the most rapid turnaround possible. We explained our reservations about the suitability of this approach in a November 3, 1981, letter.

SECOND APPROACH--CONTACTING DATA SPECIALISTS

Abandoning the matrix approach, we conferred with the Special Committee about developing a more timely approach. We first consulted knowledgeable members of the gerontological research community and compiled a list of data specialists, from which the Special Committee selected five for each of the four policy areas--economics, health, housing, and demographics. We then contacted these experts and, separately for each policy area, developed lists which summarized their recommendations on automated data files that are suitable for policy analysis.

First, we present, separately for the four policy areas, the automated files that received three or more recommendations from the data experts. In economics, all five experts recommended the Longitudinal Retirement History Survey and the National Longitudinal Surveys of Labor Market Experiences; three suggested the Current Population Survey. In health, four recommended the Health

Interview Survey; three nominated the National Nursing Home Survey, the National Medical Care Expenditure Survey, and the public use files of Medicare tapes. In housing, recommendations were diverse, with little overlap among the experts about the data; one expert did not recommend any tapes, two suggested using the facilities at the University of Michigan's National Archive of Computerized Data on Aging, and the remaining two named specific data files. In demographics, all five data specialists recommended two data sets--the public use files from the Bureau of the Census and the Current Population Survey. The Longitudinal Retirement History Survey and the Annual Housing Survey each received three votes from the demographics data specialists.

The complete enumeration of recommended data files follows. For each of three policy areas, we give two lists. The first presents data files that received more than one recommendation. After each item in this list, we note the actual number of recommendations in parentheses. The other list names data files that received only one recommendation. Only one list is possible for housing because there was no overlap between the experts.

Economics

Five data sets on economics received multiple recommendations.

1. Longitudinal Retirement History Study by the Social Security Administration (5)

2. National Longitudinal Surveys of Labor Market Experiences by the Department of Labor (5)

3. Current Population Survey by the Department of Labor (3)

4. Continuous Work History Sample by the Social Security Administration (2)

5. Panel Study of Income Dynamics by James Morgan at the University of Michigan (2)

Nine other files on economics were each recommended by one expert.

1. Census 1970 public use file by the Bureau of the Census

2. Consumer Expenditure Study of 1971-72 by the Bureau of Labor Statistics

3. CPS-SSA-IRS Exact Match File by the Social Security Administration

4. Defined Benefits Plan Characteristics by the Bureau of Labor Statistics

5. Duke Longitudinal Study by Erdman Palmore at Duke University

6. Longitudinal Employee-Employer Data (LEED) File derived
    from the Continuous Work History Sample (item 3 in the
    list above)

7. Nationwide Food Consumption Survey by the Department
    of Agriculture

8. Newly Entitled Beneficiaries by the Social Security
    Administration

9. Survey of Income and Education by the Bureau of the
    Census

Economics data specialists noted two important matters that
the Special Committee should consider. One is that the specific
characteristics of the data files should be accounted for, dif-
ferentiating whether the data are longitudinal or cross-sectional.
The other pertains to how data are categorized, some data sets
being about the economic situation of individuals while others
are about economic institutions that affect individuals. In
addition, one data specialist stated that the Special Committee
could be better assisted if the policy area was more specifically
focused.

Health

Six data files on health received multiple recommendations.

1. Health Interview Study by the National Center for Health
    Statistics (4)

2. Medicare public use files by the Health Care Financing
    Administration (3)

3. National Medical Care Expenditure Survey by the National
    Center for Health Statistics (3)

4. National Nursing Home Survey by the National Center for
    Health Statistics (3)

5. Massachusetts Health Care Panel Study by Laurence Branch
    at Harvard University (2)

6. National Medical Care Utilization and Expenditure Survey
    by the National Center for Health Statistics (2)

Fourteen data files on health were each recommended by one expert.

1. American Hospital Association provider data

2. American Medical Association provider data

3. Area Resource File by the Health Resources Administration

4. Framingham Heart Study by the National Institutes of
    Health

5. Hospital Discharge Survey by the National Center for
    Health Statistics

6. Household Pension Survey conducted for the President's
    Commission on Pension Policy and available from National
    Technical Information Service of the Department of
    Commerce

7. National Ambulatory Care Survey by the National Center
    for Health Statistics

8. National Survey of the Aging by Ethel Shanas at the
    University of Illinois at Chicago

9. Section 222 Daycare and Homemaker Demonstration Evalua-
    tion by the National Center for Health Services Research

10. Service Organizations and the Public by Saad Nagi of
    Ohio State University

11. Study of the Well-Being of Older People in Cleveland,
    Ohio, by the General Accounting Office (one expert rec-
    ommending it while another expressed reservations)

12. Survey of Income and Education by the Bureau of the
    Census

13. Survey of Institutionalized Persons by the Bureau of
    the Census

14. Survey of Low Income, Aged, and Disabled by the Social
    Security Administration

A major consideration among the health data specialists was
the sample. Several suggested that the Special Committee focus
on data files based on national samples. Another put stress on
the importance of weeding out nonrepresentative samples and con-
sidering only data tapes that are based on representative samples
from identifiable populations.

Comments were offered on three other issues. One expert
suggested using only longitudinal data sets and eliminating
cross-sectional data because cross-sectional data provide little
information on issues concerning health and aging. Another ex-
pert indicated that there is a need for data files on patients
and on providers. On the specific topic of long-term care for
the elderly, one expert recommended that the Special Committee
consider only national data files that have large samples and
that consequently can provide sufficiently large subsamples of
elderly with disabilities and chronic illnesses. One expert

also requested more specificity from the Special Committee with
respect to the particular set of health policy issues.

Housing

We found greater diversity of opinion among the housing
experts than among experts in the three other policy areas.
One data specialist would not recommend any data files as suit-
able for housing policy issues as they pertain to the elderly.
Two other experts strongly recommended that the resources avail-
able at the University of Michigan's National Archive of Compu-
terized Data on Aging be used. Only two data specialists offered
specific names of existing automated files; between them, they
recommended a total of eight data files with each file receiving
one nomination.

1. Annual Housing Survey by the Department of Housing and
    Urban Development (highly recommended by one expert,
    another indicating that it is not appropriate for the
    housing policy concerns of the elderly in the United
    States)

2. Census 1980 data file

3. Conversion of Rental Housing to Condominiums and
    Co-operatives by the Department of Housing and Urban
    Development

4. Experimental Housing Allowance Program by the Department
    of Housing and Urban Development

5. Housing for the Elderly Handicapped: Study of Section
    202 by the Department of Housing and Urban Development

6. Massachusetts Institute of Technology recent data set
    from five areas in the United States available through
    Sandra Howell and addressing policy issues specifically
    related to the housing needs of the elderly

7. Problems in Multifamily Housing Projects by the Depart-
    ment of Housing and Urban Development

8. Survey on the Quality of Community Life by the Depart-
    ment of Housing and Urban Development

Demographics

Six files on demographics received multiple recommendations.

1. Bureau of the Census tapes (5) (experts having recom-
    mended particular favorites including 1910, 1940, 1950,
    1960, 1970, and 1980)

2. Current Population Survey by the Department of Labor (5)
    (some experts suggesting specific months, such as March)

3. Annual Housing Survey by the Department of Housing and
    Urban Development (3)

4. Longitudinal Retirement History Study by the Social
    Security Administration (3)

5. Health Interview Survey by the National Center for
    Health Statistics (2)

6. National Center for Health Statistics micro data on
    multiple causes of death (2)

Eight additional files on demographics received one recommenda-
tion.

1. Income Survey Development Program by the Social Secu-
    rity Administration

2. Longitudinal Study of Housing Adjustments Among the
    Elderly by the Department of Housing and Urban Devel-
    opment (pretested but not fully implemented)

3. National Crime Panel Victimization Survey by the Depart-
    ment of Justice

4. National Nursing Home Survey by the National Center for
    Health Statistics

5. Newly Entitled Beneficiaries by the Social Security
    Administration

6. Social Security Administration data file containing
    Social Security and Supplemental Security Income and
    also ZIP code data

7. Survey of Income and Program Participation by the Social
    Security Administration

8. Survey of Institutionalized Persons by the Bureau of
    the Census (one expert recommending it while another
    expressed serious reservations about the adequacy of
    its documentation)

Experts raised three important points to keep in mind about
demographic data sets that might be suitable for policy purposes--
that the national samples should be adequate, that the reliability
of data should be known, and that the data should be robust. In
addition, one wanted to lay stress on the fact that a number of
analytic issues must be considered in the secondary analysis of

an existing data file, including the intention behind the original
survey, the characteristics of the survey nonrespondents, and the
way survey responses were coded.