DESIGNING AN INTEGRATED DISABILITY DATA SYSTEM FROM SOCIAL SECURITY ADMINISTRATIVE RECORDS

Robert E. Klein, Social Security Administration
Daniel Kasprzyk, Bureau of the Census

INTRODUCTION

The Social Security Administration's Title II disability insurance program (DI) is over twenty-five years old. Since its enactment, the program has gone through major expansion along with considerable administrative changes. For example, in 1960 there were about one-half million disabled worker beneficiaries [1]. But by 1978 the rolls peaked at 1.9 million. The secular increase in the number of beneficiaries is explained, in part, by the increase in the absolute number of workers insured under the program. But we also observe a secular increase in the rate of insured workers receiving benefits, i.e. beneficiaries per 1,000 insured workers. Although of late there has been a dip in this rate, there is no attendant dip in the monies distributed. Between 1977 and 1981, cash paid out increased from $11.5 billion to $17.2 billion, in part because benefits are tied to the Consumer Price Index [2].

In the past year or so, much attention has been focused on both Title II DI and Title XVI SSI (SSA's other major disability program) by both the government and by the public through the national news media. The reason for this attention centers primarily on the public's perception of the consequences of changes in the law enacted in 1980. The 1980 Amendments set out not only to enhance work incentives in the disability programs but also to improve the administration of the programs to insure that benefits go to and are continued for those who are genuinely eligible [3]. The provisions to implement these goals include a requirement that all nonpermanent disability cases be reviewed at least every three years. This is the Continuing Disability Investigation, or CDI. More CDIs, coupled with other administrative changes, have resulted in an increase in the number of decisions to discontinue benefits [4].

Both the decline in initial awards and the imposition of seemingly more stringent criteria for continuing disability benefits have sparked the interest in the disability programs. This interest coupled with a) the sequential nature of the application, review and appeals procedures, b) the long duration of disability for many beneficiaries and c) the reapplication and appeal of many people denied or ceased benefits suggest that on-line longitudinal files are necessary to provide an accurate account of cases. With this in mind, we will explore current data bases from disability program administrative records, indicate longitudinal capabilities, and present a relatively simple approach to creating a new data system by paring down existing data bases and synthesizing them. To document the sources of the various existing data bases, we first outline the procedures for initial disability claims and for CDI, since data are generated at each step in the process.

INITIAL DISABILITY CLAIMS

A claim for Title II or Title XVI disability benefits starts with filing an application in a district office (DO). The DO makes an initial determination of eligibility by establishing that: a) the claimant is not engaged in what is called "substantial gainful activity" (SGA) (i.e. not earning more than $300 per month), b) the claimant has worked enough quarters to be insured under the program and, additionally for SSI applicants, c) a means test is met. Once eligibility on nonmedical grounds is established, the application is sent to the State's Disability Determination Service (DDS) for review of medical evidence. The DDS determines if supporting medical evidence indicates a) that the impairment is expected to last (or has lasted) at least 12 months, or will result in death, and b) that the evidence shows the impairment equals or meets the degree of severity listed in the regulations. If the medical evidence indicates that an impairment does not meet or equal the degree of severity listed, an assessment is made of the claimant's residual functional capacity, that is the claimant's ability to perform relevant previous work. If past work can be performed, the claim is denied. If the claimant is unable to do previous work, nonmedical factors such as age, education, training and work experience will be evaluated in the context of the medical evidence to determine whether the claimant can do other work found in the economy.

Claimants denied disability benefits under DI or SSI (or both) by the DDS have legal recourse to appeal the decision. There are several distinct stages for such an appeal: (1) reconsideration by the DDS; (2) a hearing before an administrative law judge; (3) review before an Appeals Council; and finally, (4) appeal in a federal district court [5]. Denials can be overturned at any step in the appeals process. In fact, a majority of claims denied by the DDS are reversed at a subsequent appeal stage [6].

CONTINUING DISABILITY INVESTIGATION (CDI)

Medical recovery or improvement need not be firmly established to determine that a beneficiary's disability has ended. It is only necessary to establish that the definition of disability is not currently met. Nonpermanent disability cases are reviewed routinely for medical evidence of disability at least every three years. Intervening events also can result in a termination. For example, program experience has identified certain impairments which may be expected to improve. Claims based on these types of impairments are flagged for medical reexamination within 12 to 24 months of allowance. Voluntary reports of work or medical recovery, as well as substantial reported earnings (which are routinely flagged), can result in discontinuation of benefits. If disability benefits are terminated, a beneficiary has the same recourse for appeal as a claimant denied benefits on an initial claim.

Administrative records on the claimant (or beneficiary) are kept for both initial applications and CDIs. The sequential and decentralized nature of the process under different organizational components highlights the complexity of tracking a case and of maintaining data files for research purposes. This paper stems from the concern that anticipated requirements for information detailing the post-adjudicative history of disability beneficiaries on the rolls cannot be met in a wholly satisfactory way with current data bases.

EXISTING DISABILITY ADMINISTRATIVE DATA BASES

Most of SSA's extant administrative record files on the disability programs exist for case control and are not designed for research purposes, per se. The information available on the files is generally limited to basic demographic data, to those data needed to make a disability determination, and information generated from the appeals process. Data on disability determination include specific and quite reliable diagnostic information [7]. While data are available on most aspects involved in the disability program, no one data base on the entire process is wholly satisfactory. If these files are pared and synthesized, an integrated data base can be created not only to meet administrative needs but also to provide a rich source of information for disability research. In this section, we describe the building blocks of an integrated data system, namely, existing files from administrative data.

1) The SSA-831-U5 File:

This file is compiled on an annual basis from all applicants for Title II and Title XVI disability benefits. The information comes from form SSA-831-U5, which is filled out in the district office by a claims representative, in part, from information provided by the claimant on the application for benefits. The 831 file is a critical file because it contains the basic personal and program data about the claim. The 100% 831 file is the universe file containing all applicants for disability benefits under Title II, Title XVI, and those applicants filing concurrently under both Titles. The information on the file includes: name; date of birth; SSN; DDS and district office; prior action (termination or denial for a previous claim); Title II, Title XVI or concurrent claim; and current determination of claim (filled out at the DDS). Most of the items on the 831 form are pre-coded, with some important exceptions. Specifically, diagnosis, industry and occupation, and mobility status are blocked on the form, but are not pre-coded. The information is eventually coded for a sample of successful applicants and denials. [See item on the Continuous Disability History Sample (CDHS).] The full 831 file has been used primarily as a frame for sampling cases for the CDHS. The 831 file may be matched with other files on an ad hoc basis for special purposes, often on a sample basis. For example, there have been plans to match 831 information with information from the Supplemental Security Record for a sample of 1977 allowances and denials. The file is expected to permit analysts to follow the cohort of applicants in terms of program experience and subsequent employment over time.

2) The SSA-833-U5 file:

The 833 file is based on the 833 administrative form (cessation or continuance of disability), filled out by the appropriate local SSA agency in a continuing disability investigation (CDI). CDIs have become more frequent as a result of the 1980 Amendments, which require mandatory triennial review of nonpermanent disabilities. This file, containing basic demographic information, is noteworthy for its information on "reason for the CDI" and "basis for the determination." It is crucial to any development of an integrated data system because when matched with basic 831 information, it represents the first step to establish a longitudinal link between initial determination and the decision to continue or cease benefits after time on the rolls.

The full 833 file also does not have completely pre-coded information blocked on the form. As is the case for the 831, the most notable datum is diagnosis. In most instances, diagnosis on the 833 file should be identical to diagnosis on the 831 since the CDI is made on the basis of the original impairment. However, intervening conditions can exacerbate an original diagnosis and/or a nonrelated ailment between the initial award and the CDI and can affect the medical determination for continuing benefits.

SSA's Office of Data Systems combines data from several sources to: a) report frequency counts on different aspects of the disability program for management purposes and b) provide an information system alerting appropriate operating components that CDIs are due. With respect to the latter purpose, the 833 file is the major component of satellite files which serve as disability master files for flags needed by the Office of Disability Operations (ODO) for periodic review of nonpermanent impairment disability cases and for the control of current pending disability cases. Other output files, which use 833 as a component, are used to create statistical reports for annual publications.

3) The Continuing Disability Investigation Management Information System (The CDI MIS):

The CDI MIS is being developed in cooperation with several SSA components to provide a data base system on most aspects of CDI activities. Its purpose, to track CDI cases and provide descriptive statistical data, is a result of the importance of the CDI process to SSA's workload in recent years. The legislation requiring periodic review as well as the establishment of improved performance standards for the DDSs increased the need for CDI information. Furthermore, these data are intended to help the SSA respond to other provisions of the 1980 Amendments which require measuring the cost-effectiveness of various CDI-related projects and evaluating decision standards, such as those used by administrative law judges in their review of appeals.

The major impetus for the CDI MIS in terms of specific data requests/requirements came from two sources. One source represents an important constituency--the local units, i.e., district offices and DDSs. The other is the SSA component which has direct responsibility for CDI policy formulation and, as such, needs timely information on all aspects of the CDI process. The CDI MIS is planned to provide the local units with the capability of accessing the computer system to obtain basic CDI information specific to the

local unit. This is the so-called GETCDI system. Basically descriptive information is thought to be necessary for the expeditious handling of CDI cases.

The basic CDI MIS is similar in form to the system we might propose. It falls short, however, in some important ways. Specifically, CDI MIS concentrates on the CDI and does not give high priority to the 831-833 merge. One extremely important aspect of this merge is the combination of personal characteristics with program information about applicants and beneficiaries. In effect, diagnosis data would be matched for individuals with personal characteristics such as age, sex, education and occupation, and with CDI program information.

4) The Continuous Disability History Sample (CDHS):

Aside from the potential CDI MIS, the CDHS comes closest to the basis for an integrated data base. The CDHS file's ostensible purpose is to furnish information and to support research studies on the Social Security Title II Disability Program. The file contains a sample of persons allowed or denied disability benefits during a calendar year. The sample represents 20 percent of allowances and 10 percent of denials. It includes demographic and diagnostic information extracted from the claims folder of each person in the sample. The data are augmented by adding benefit and earnings information from other administrative record operations--specifically, the Master Beneficiary Record and the Summary Earnings Record. The CDHS is cross-sectional; it is intended for longitudinal analyses. But it is neither continuous nor historical, and would be greatly enhanced if prospective events were tracked for the annual cohort of applicants. The "historical" aspects of the information contained in the CDHS are limited to retrospective information contained in the Title II benefit payment records. Furthermore, the file's utility would be enhanced if it were more timely. At the moment, a file for a cohort of applicants in a given year generally becomes available in not less than three years. Published results lag from the time the data are collected by about five or six years. The utility of such cross-sectional data diminishes rapidly as the data age, particularly in light of changes brought on by the enactment of the 1980 Amendments. At this writing, a cross-sectional file for 1979 is being prepared; 1976 is the last date for which CDHS tabulations have been published [8].

5) The Master Beneficiary Record (MBR):

The MBR is a major source of Title II disability information because it triggers payments to beneficiaries. The MBR consists of the universe of claimants for Title II benefits: past and present cash beneficiaries, including disability claimants; disallowed claimants (i.e. nondisability aspects of claims); denied claims on disability, and concurrent claims for disability benefits under Titles II and XVI. Denials under Title XVI are excluded. The MBR also contains information beyond that of immediate relevance to current payment. For example, the file may contain up to six fields of records providing a history of a beneficiary on the rolls. While the main purpose of the MBR is to trigger payments and catalogue benefit information for operational purposes, it is used frequently by researchers because it is updated regularly and has extensive program-related information and basic demographic data. In addition, it is often linked with other files--the CDHS is a prime example.

6) Hearings and Appeals Information:

Considerable data are generated by the Office of Hearings and Appeals (OHA). These data have been used recently to comply with those sections of the 1980 Amendments which deal with the appeals process. The major purpose of the Office's data gathering activity is to maintain a sophisticated case control system.

As it is presently constituted, the system receives information through on-line terminals from the regional units. Each action which occurs within a Hearing Office is coded and sent to the appropriate regional unit. Examples of actions taken within the Office are: assigning an Administrative Law Judge, scheduling a hearing, recording the decision, and sending favorable decision information to the Payment Center and the District Office. Each action in SSA headquarters is processed so that current information is on-line for both local units and headquarters. Every three months on-line data are purged of records over seven months old. Cases remain in the Management Information files for 13 months. While the system which generates the data is mainly used for case control, several monthly and quarterly reports are produced from these data on case load and other management information for local units and headquarters.

7) Supplemental Security Record:

The Supplemental Security Record (SSR) is the basic tape file for data on eligibility, payment amounts and characteristics of all Title XVI (SSI) beneficiaries and applicants. The SSR contains the records of 9.4 million persons. These include persons currently receiving federally administered SSI payments, those who received payments in the past but are not newly-eligible for benefits, and persons whose applications for payments were denied. In April 1978, records in denial status for a year or longer were removed from the SSR. These are carried in a separate accessible file in the same format as the SSR.

The SSR is an important disability data source because it has information on individuals who apply for benefits under Title XVI. It would figure prominently in any plan for an integrated disability file, particularly in light of the proposed longitudinal aspects of such a file. The retrospective portion of each record on the SSR records the Federal SSI and State supplementation amount of every benefit check authorized. Currently, this information is carried from January 1974 to date.

DESIGN FOR THE DATA SYSTEM

Current thinking on the design for an integrated disability data system is the culmination of preliminary efforts to: a) ascertain which files exist, b) assess the strengths and weaknesses of current files, c) explore data needs within SSA, and d) evaluate how current capabilities match up with current needs. The resulting blueprint represents a preliminary estimate of how unmet needs can be met. Those unmet needs generally center on following the post-entitlement history of beneficiaries, monitoring reapplication of those denied benefits or terminated from the rolls, and linking program characteristics of beneficiaries with med-

ical and sociodemographic variables. The scheme proposed is a first step, on paper, to build such a longitudinal file. The next step, and perhaps a far more difficult task, is undertaking a feasibility study to test the viability of such a scheme.

Current thinking on the integrated data system can be divided into four phases for the current effort and a final phase to consider longer range issues. These four phases are:

I. Deciding on a universe or universes. What are the appropriate kinds of cases from which to make a selection for a file sample--e.g. all DI applicants under a given Title; awards, denials, etc.? We need to define a universe [and then match the universe with the appropriate file(s)].

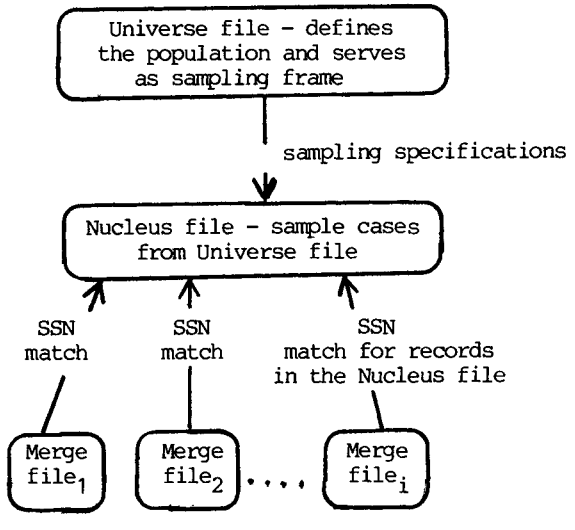
II. Selecting a sampling frame which defines the universe in a statistical sense, since cases will be randomly selected from such a file to serve as the nucleus file on which to build an integrated system.

III. Determining sample(s) specifications. This includes determining the nature of the sample(s)--simply random sample or stratified random sample--and the size of the sample.

IV. Determining file content. This involves deciding what the general content areas of the file should be, and what specific data elements to include in a basic integrated file.

Each phase is an integral part of general strategy that involves using the universe file to construct a nucleus file by sampling cases from the universe and merging information on those sample cases with other files to construct a file that is as integrated and longitudinal as possible. Using the Social Security Number (SSN), we would merge information from other files to data elements contained in the nucleus files. The strategy is basically very simple and is represented schematically in Figure 1.

Figure 1--The Basic Design

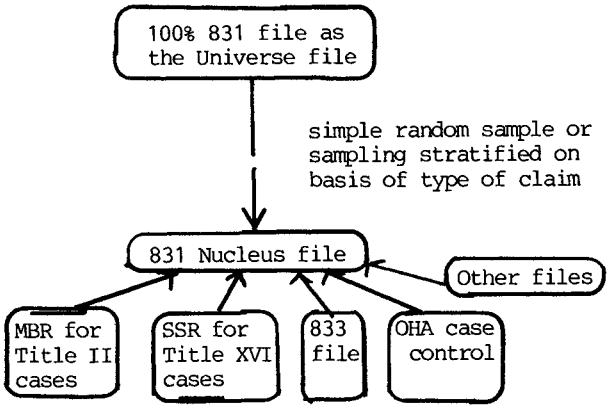


There are four populations of concern to us. The first three are defined by initial applicants in a given year for disability insurance benefits under 1) Title II, 2) Title XVI, and 3) both Title II and Title XVI, concurrently. A fourth population is defined by the Continuing Disability Investigation (CDI) cases in a given year--i.e., what we will call "current" CDI cases.

What then, are the universe files which cover these populations? The choice of a universe file depends, in large measure, on the issues determined by potential users to be of greatest importance. For the first three populations, i.e.; initial DI applicants--applicants under Title II, Title XVI and "concurrent applicants"--there are two choices for universe files from which to draw sample cases or, more precisely, there are two design strategies which speak to the first three populations.

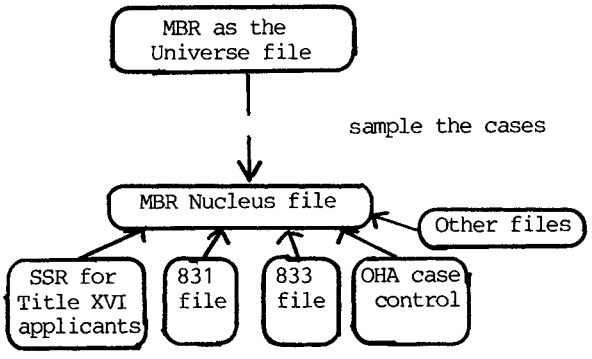
One choice is to use the 100% 831 file as a universe file. The SSA 831 form, which is the initial application determination form, contains a block indicating Title under which the claim is made as well as the determination. The full 831 file is used currently as a frame from which records are sampled for the CDHS. Since it includes application for benefits under both Titles and since the initial determination is indicated on the form, this file seems to be a logical choice for a frame from which to draw a sample for a new integrated disability data system. It has the advantage of including cases under both Titles. However, essential impairment-related information, such as mobility, and diagnosis, must be hand-coded, as is presently done with the CDHS. Other information, related to personal characteristics, such as occupation and industry, must also be hand-coded and processed for the file. If the 831 file is used, denials as well as awards are readily available for separate files. Since the 831 file contains duplicate listings, individuals, if denied, may apply again. Therefore, if interest centers on persons applying for benefits, duplications would have to be accounted for in the file. This has implications for following denied applicants longitudinally. If the full 831 file is used, cases can be randomly selected on a simple random sample basis or on a stratified basis, using the type of claim (Title II, Title XVI or concurrent claims) as a stratifying criterion. The nature of the sample depends on the size of sample selected and the relative distribution of applicants under each type of claim in the universe of applicants. We can have a sample of denials as well as awards. Once cases are selected from the 100% 831 file to form an 831 nucleus file, information from other files can be merged with the 831 nucleus on an ongoing, updated basis. These other files include: MBR information to supplement Title II information for Title II applicants; SSR information for Title XVI applicants; 833 information for post-entitlement events for awardees; and other files, such as hearings and appeals (OHA) data from their case control system, Office of Assessment's Quality Assurance files, and the CDI MIS. This first alternative to address the population of initial applicants is mapped out in Figure 2, which follows. This scheme has the advantages attending the 831 form--it contains much information on all the types of claims. It uses the CDHS as a model. Its disadvantages result from the need to hand-code much information and the fact that some of the current CDHS problems may remain.

Figure 2--Design with 831 File as Universe File



The second choice for a universe file for the population of initial applicants is the MBR. The MBR defines the universe of applicants under Title II (awards and denials) and flags Title XVI awards. It does not have information on Title XVI denials. If cases from the MBR as a universe file are randomly selected, the MBR information could be supplemented with records from other files, such as the 831 file, the SSR for Title XVI awards, the 833 for CDI and others. The scheme can be seen in Figure 3.

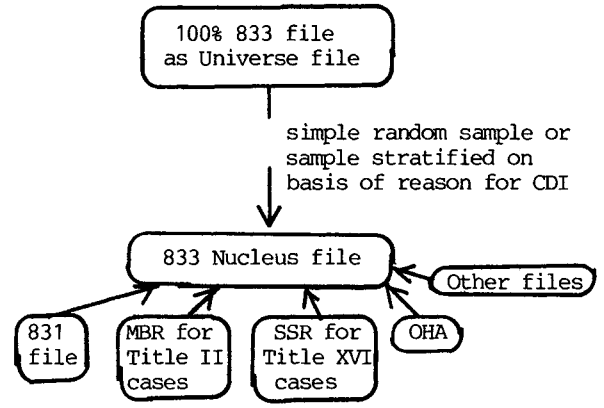
Figure 3--Design with the MBR as Universe File



The advantage of the MBR as a universe file over the 831 file is that the MBR is ongoing and thoroughly established within the data processing components of SSA. It now has diagnosis coded as a matter of course. Its major disadvantage is that the data only pertain to Title II and concurrent beneficiaries, although we can identify Title XVI awardees from the MBR. Detailed information on Title XVI recipients and Title XVI denials would have to be obtained from the Supplemental Security Record (SSR). Since only concurrent recipiency of benefits is flagged on the MBR, details on Title XVI recipiency for concurrent beneficiaries would have to be obtained from the SSR. Furthermore, the MBR does not have fields for occupation, industry and other impairment-related information, which is blocked on the SSA 831.

Data on the CDI are of great interest within SSA. If a cohort of initial beneficiaries is followed longitudinally, we will have data on CDI. However, we need to address issues concerning current CDI cases. Therefore, a separate file of CDI cases must be developed from a suitable frame. This is the fourth population defined earlier. The 100% 833 file can serve as a frame from which to sample CDI cases, either on a simple random sample basis or stratified on the basis of the reason for the CDI. Once cases are selected from the 833 file, information on these cases can be matched with case records from other files, including 831 information on the initial application. This scheme is shown in Figure 4.

Figure 4--Design with 833 File as Universe File



The content of these files should center around three general areas of substantive interest, namely:

1. individual, sociodemographic and socioeconomic characteristics, such as age, sex, race, occupation and industry. These data, linked with information in the other general areas, will be of the greatest value to SSA researchers. However, other components have expressed their interest in such a link.
2. program-related characteristics, such as level of adjudication and reason for CDI.
3. impairment-related characteristics, such as diagnosis, mobility, body system, severity, and "meets" or "equals" specific medical criteria.

The content of the files should contain basic elements within these three general areas. The files should be kept "basic" to expedite their processing. Any additional, one-time needs can be met through satellite or extract files.

Another important issue is the periodicity of the file. For example, since it is desirable to follow a cohort of applicants awarded benefits while they are on the rolls, should a new annual cohort of awardees be tracked each year? For research purposes, it is probably not necessary nor practicable to devote resources to a longitudinal file for each annual cohort of applicants. However, those within SSA who need management information files might feel that annual files are necessary.

SUMMARY

The process by which an applicant is determined to be eligible for Social Security disability benefits, on both medical and nonmedical grounds,

is sequential, multi-staged and complex. If an applicant is denied benefits at the initial stage, several formal channels of appeal are available to reverse the decision. Administrative data on program-related matters, medical condition, and personal, sociodemographic characteristics of the applicant are collected at each stage. While a plethora of data exist, these data are segmented. Data following all stages of this process, as well as post-adjudicative events, are not linked currently in a longitudinal file, although building a longitudinal disability data base is a relatively simple idea to put on paper. It involves sampling cases from a universe file and linking the independent files for the different stages or aspects of the process. Implementation is another matter, however. Issues of file content as well as current updating capabilities must be addressed. More important, a great commitment of both data processing and human resources are required for a viable integrated data system.

ACKNOWLEDGEMENTS

We thank Emily Andrews for many thoughtful comments and helpful suggestions and Faye Aziz for her work on the project. Gratitude is extended to Marie Colman, Weltha Logan and Helen Kearney for their help in preparing the manuscript. Any remaining shortcomings of this paper are our responsibility.

REFERENCES

[1] Social Security Administration. 1982. Annual Statistical Supplement, 1981. Social Security Bulletin, Vol. 45, No. 11 (November), p. 105.

[2] Lando, Mordechai E., Alice V. Farley and Mary A. Brown. 1982. "Recent Trends in the Social Security Disability Insurance Program," Social Security Bulletin, Vol. 45, No. 8 (August), pp. 4-5.

[3] U.S. Senate, Special Committee on Aging. 1982. Social Security Disability: Past, Present and Future. Washington: U.S. GPO, p. 17.

[4] U.S. House of Representatives, Subcommittee on Social Security. 1981. Status of the Disability Insurance Program. Washington: U.S. GPO, p. 9.

[5] Social Security Administration. 1982. "The Bellmon Report," Social Security Bulletin, Vol. 45, No. 5 (May), pp. 7-8.

[6] Reference [3], p. 2.

[7] Krute, Aaron, Barry Bye, and Audrey Coe. 1978. "SSA Public-Use Data Files for Disability Research." Pp. 104-116 in Policy Analysis with Social Security Research Files. Washington: U.S. DHHS, SSA, p. 111.

[8] Burdette, Mary Ellen and Shirley Baker. 1982. Characteristics of Social Security Disability Insurance Beneficiaries, 1976. Washington: U.S. DHHS, SSA.