

CHAPTER 5

CORSIM

I. OVERVIEW	5-1
II. BACKGROUND	5-1
III. DESCRIPTION	5-2
Components	5-2
Controls and Alignment	5-17
Databases	5-18
Documentation	5-19
Computer Hardware, Software and Portability	5-19
IV. APPLICATION TO IMPORTANT RETIREMENT POLICY ISSUES	5-20
Benchmarks	5-20
V. ACCESSIBILITY AND EASE OF USE	5-22
VI. CRITIQUE	5-23

ANNEX 5-1 -- CORSIM 3.0 SUMMARY DESCRIPTION TABLE

ANNEX 5-2 -- CORSIM POLICY MATRICES

ANNEX 5-3 -- PROPOSED CORSIM PRIVATE PENSION MODULE FLOWCHARTS

CHAPTER 5

CORSIM¹

I. OVERVIEW

CORSIM (Cornell Microsimulation Model) is a large dynamic microsimulation model developed and maintained at Cornell University. It currently ages the 1960 Census Public Use Microdata Sample (PUMS), operating on a sample of up to 180,000 person records. It ages the database in annual periods from 1960 through 2030. The model depicts most of the important demographic events and characteristics, but has minimal economic behavioral content. It has a detailed social security module, with comprehensive representation of social security rules, but little behavioral analysis. CORSIM staff have designed a private pension module, but it has not been developed. CORSIM does not include industry affiliation of workers. It does not depict behavior or characteristics of employers or retirement plan sponsors.

II. BACKGROUND

Development of CORSIM began at Cornell University in 1986 under the direction of Prof. Steven Caldwell. Caldwell was part of the team that developed DYNASIM at the Urban Institute in the early 1970s (Orcutt *et al.*, 1976).² The first version of CORSIM was similar to DYNASIM. A second version was developed during 1990-1993 with funding from the National Institute of Dental Research, which included a detailed dental condition and cost module. In 1994 and 1995 the model was broadened, with support from the U.S. Department of Health and Human Services, the Bureau of the Census, the Internal Revenue Service, the Canadian government Office of the Superintendent of Financial Institutions, and the American Association of Retired Persons. The version current since 1995 is termed CORSIM 3.0.

In 1995 CORSIM was adopted by the Office of the Superintendent of Financial Institutions (OSFI) of the Government of Canada to serve as the basis for the development of a microsimulation model to be used for analysis of Canada Pension Plan policy. The OSFI has revised and replaced

¹ This description draws on Caldwell (1993), Caldwell (1995), Caldwell *et al.* (1997), Caldwell and Keister (1995), Caldwell and Favreault (1995), Williams (1997) and discussions with Steven B. Caldwell, Melissa Favreault, and Thomas Johnson of Cornell University.

² DYNASIM is described in Chapter 3 of this report.

CORSIM modules to develop a model of the Canadian population and retirement income system, called DYNACAN.³

The primary objective of CORSIM is to provide a framework for research concerning the socioeconomic behavior and states of the population, particularly life-course patterns and outcomes. A second objective is to provide a platform and framework for research in microsimulation modeling techniques, for example investigation of computer time requirements for model simulations, alternative variance reduction techniques (discussed below), and use of alignment factors for simulation.

A key objective for the CORSIM model has been portability among computational platforms. An important aspect of this was to develop a model that could be run with small to moderate size samples on a desktop micro computer and ported to a large mainframe for large samples.

III. DESCRIPTION⁴

CORSIM is a large-scale dynamic microsimulation model.⁵ It simulates the basic demographic and socioeconomic events that characterize the socioeconomic conditions of the population. The basic elements of CORSIM are persons and families. It does not depict firms, industries, or governmental entities. It simulates the basic demographic and socioeconomic events, states, and changes occurring for each person and family in its base microdata file on an annual basis. It depicts approximately 26 events, processes, or characteristics of persons and families and can operate on a database of up to 180,000 records. The model is usually run over the period 1960-2030, but for social security simulations it has been run for 1960-2090.

Components

CORSIM is composed of a single integrated simulation model that operates on a microdata file, and a set of tabulation and report producing programs that can be used to tabulate, analyze, report, and display the simulated output data. The structure of CORSIM is shown in Figure 5-1.

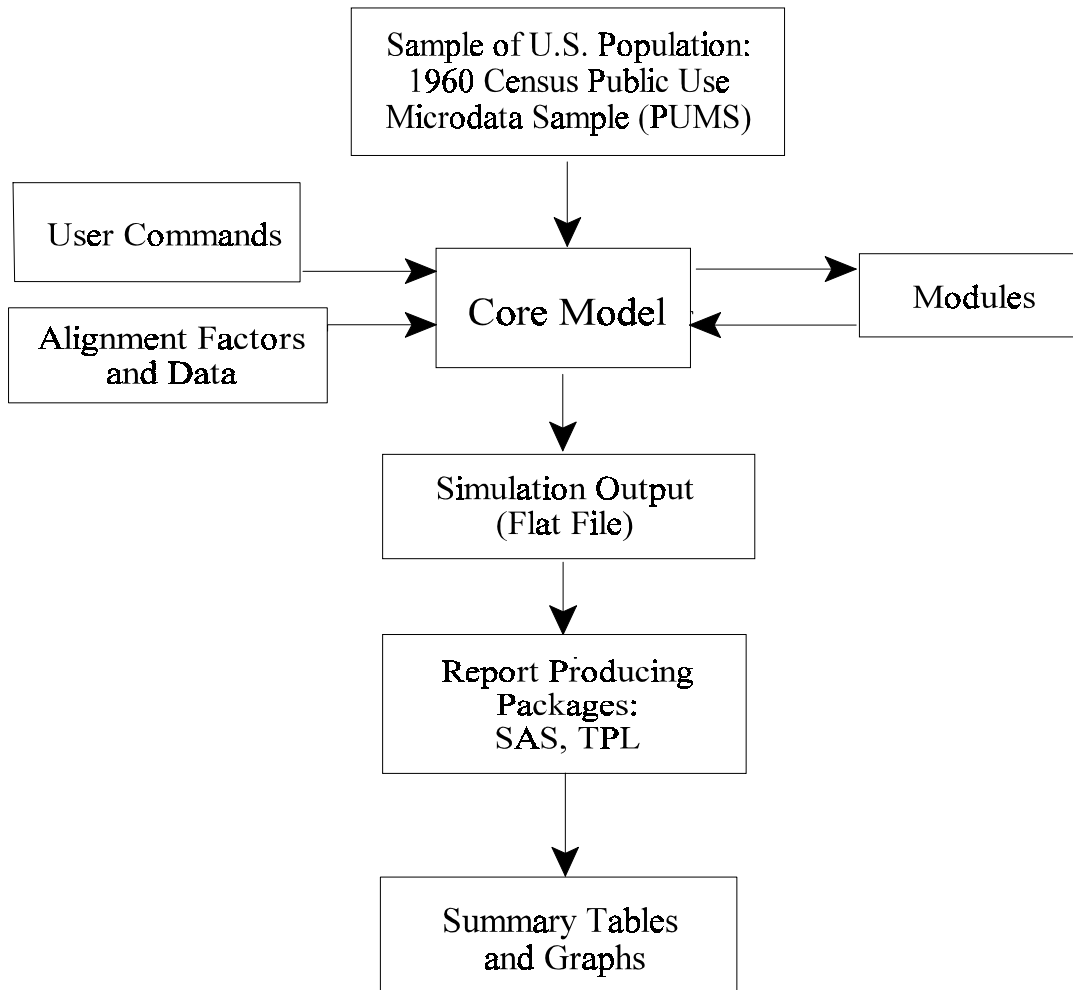
³ DYNACAN is described in Chapter 6 of this report.

⁴ The basic features of CORSIM, including sources of data, are summarized in the Summary Description Table in Annex 5-1 at the end of this chapter.

⁵ Dynamic and static microsimulation are described in Chapter 2. As described in Chapter 2, a microsimulation model simulates social and economic behavior by depicting events, conditions, and changes in the information recorded on each individual (person or family) record in a large database. A microsimulation model depicts the aggregate conditions of the population by aggregating or tabulating over all the modified individual records. Dynamic means the model simulates events for each individual one year at a time, and the outcome for each event each year depends on current and past year's outcomes for that and other events.

FIGURE 5-1

STRUCTURE OF CORSIM



CORSIM is fully dynamic. Each event, characteristic, or process for each person or family is determined for each year, one year at a time, with reference to variables representing demographic and socioeconomic characteristics of the person or family in that year and previous years, including lagged (previous years) values of the variable being determined, and variables characterizing the socioeconomic environment.⁶ For example, annual weeks worked for each person is determined by an equation including as explanatory variables demographic characteristics of the person (e.g., age,

⁶ Family wealth is simulated at three year intervals only.

sex, race) and the number of weeks worked in the previous year.⁷ Each of the behaviors, events, or processes depicted in CORSIM is included in an individual module. A module may determine outcomes using rule-based processes or behavioral equations or both. Table 5-1 shows the events and processes depicted in CORSIM and the variables used to determine each event. Not every module or process is necessarily simulated in every run of the model. The order in which the modules are solved is shown in Table 5-2.

CORSIM simulates all of the appropriate events for each individual in the file each year.⁸ In the versions of CORSIM existing through 1997, all of the events/processes are simulated sequentially for the first individual, then all are simulated for the second individual, etc., until all individuals and families have been processed for one simulation year. Then the model repeats the process for the next year, with the simulated event values for the previous year and those already determined for the current year as inputs. The process is repeated for each year until the final year of the simulation period. The output of the model is a file of longitudinal demographic and socioeconomic histories for each person in the sample of the U.S. population, including weeks worked and earnings.

In 1997 the events/persons solution sequence in CORSIM is being changed. In the revised version, for each year the first event/process is determined for all individuals, then the next event is determined for all individuals, until all events and outcomes have been determined for all individuals and families in that year. Then the model repeats this process for the next year. This change will not change the order in which event/process modules are solved, nor the form of the output database.⁹

CORSIM includes approximately 26 behavioral modules and several rule-based or bookkeeping modules. Not all the modules are included in every simulation of the model. In particular, the modules addressing behavioral risks (smoking, alcohol, sugar, diabetes), Aid to Families with Dependent Children (AFDC), Supplementary Security Income (SSI), Earned Income Tax Credit (EITC), the Social Security module and the wealth module may not be included in every simulation. In addition, three modules are separate from the main model: the voting module, the consumption expenditures module, and the dental module. These accept outputs from other modules in the model but do not feed back into other modules.

⁷ The use of the lagged value of an attribute (dependent variable) as an explanatory variable for that attribute is sometimes referred to in the microsimulation literature as state-dependence. State-dependence is a key issue in dynamic aging. In fact, relatively few events or characteristics in CORSIM are determined by lagged values of the dependent variable (see Table 5-1).

⁸ Family wealth is simulated every three years.

⁹ The solution sequence in use through 1997 is sometimes referred to as “events within persons”. The post-1997 sequence is referred to as “persons within events”. This change is essentially a programming change and will not change the outputs or most capabilities of the model. It will facilitate the use of particular alignment techniques (discussed below).

TABLE 5-1

DETERMINANTS OF MAJOR EVENTS SIMULATED BY CORSIM

Event or Characteristic	Variables Used to Determine Event ^a
(1) <i>Family demographic events</i> ^b	
Death	Age, race, sex, marital status, education, family income, employment status, birth place (from U.S., other), year
Birth	Age, race, marital status, duration of marriage, birth(t-1), birth(t-2), number of children, education, school attendance, employment, living arrangements, family income, earnings, home ownership, work status (full-/part-time)
Sex of newborn	Race
Multiple Births	Race
First marriage	Age, cohort, race, sex, education, school attendance, weeks worked, earnings, number of children
Remarriage	Age, race, sex, previous marital status (divorced/widowed), income, number of children, weeks worked, education
Mate matching	Difference in age, difference in education, difference in income, race match, number of children, labor force status, employment, income, state of residence
Leaving family of origin ^c	Age, race, sex, school attendance, education, number of parents, parents' education, employment, earnings, presence of children, lagged living arrangements, sibling age
Divorce	Age, cohort, race, age difference, duration of marriage, presence of children, wages, wife's earnings

TABLE 5-1 (continued)

DETERMINANTS OF MAJOR EVENTS SIMULATED BY CORSIM

Event or Characteristic	Variables Used to Determine Event ^a
Custody of children at death of parents	Age, state of residence
Custody of children at divorce of parents	Age of child, biological ties, if equal ties random assignment of 90 percent to mother
Immigration	Age, race, sex, marital status
(2) <i>Education, location, disability</i>	
Education (grade completed)	Race, sex, age, parents' education, living arrangements, parents own home, marital status, has child
Residential location and migration	State of residential location, characteristics of family (under development)
Disability	Disability status in previous period, age, race, sex, work status, education, family status, income
(3) <i>Labor force events and earnings</i>	
Annual weeks worked	Age, race, sex, marital status, presence of children, lagged weeks worked (0, 1-47, >47), education, number of children, age of youngest child, unemployment rate (estimated but not currently activated)
Wage (per week)	Age, race, sex, marital status, presence of children, weeks worked, education, number of children, age of youngest child
Income from employment	Product of wage times weeks worked
Labor force participation	not explicitly determined in CORSIM
Hours in the labor force	not explicitly determined
Unemployment	not explicitly determined
Employment	not explicitly determined
Industry	not determined

TABLE 5-1 (continued)

DETERMINANTS OF MAJOR EVENTS SIMULATED BY CORSIM

Event or Characteristic	Variables Used to Determine Event ^a
<i>(4) Income from transfer payments</i>	
Supplemental Security Income (SSI) (aged, disabled)	
Eligibility	State-/ marital status-specific means test, income, assets
Participation, benefit receipt	Race, marital status, earnings, asset income, education, live alone, size of OASDI benefits, work hours
Aid to Families with Dependent Children (AFDC)	Age, race, marital status, previous AFDC status and experience, duration of AFDC receipt, months ever on AFDC, time since last AFDC spell, work status, work status lagged, age youngest child, disability status, education, region, size of benefit, unemployment rate, year (AFDC module not always operational)
Earned Income Tax Credit (EITC)	Earnings, number of children, marital status, sex
<i>(5) Social security eligibility and benefits</i>	
OASDI coverage	Randomly assigned based on aggregate coverage ratio
OASDI contributions	Social security coverage, earnings, tax rules, year
Primary Insurance Amount (PIA)	Age, calendar year, coverage and earnings history, program rules
Retirement (OAI) benefit eligibility	Age, covered earnings
Retirement benefit receipt/timing	Age, race, sex, coverage, work hours (weeks), change in work hours, earnings, asset income, change in income, education, home-ownership, live alone, marital status, marital disruption
Retirement benefit level	PIA, coverage and earnings history of both worker and dependent, current earnings, earnings test rules, retirement timing, year
Disability (DI) benefit eligibility	Age, covered earnings

TABLE 5-1 (continued)**DETERMINANTS OF MAJOR EVENTS SIMULATED BY CORSIM**

Event or Characteristic	Variables Used to Determine Event ^a
Disability benefit receipt	Age, race, sex, coverage, disability status, receipt of DI benefit in previous period, work hours in previous period, change in work hours, earnings, asset income, income, education, home-ownership, live alone, marital status, marital disruption
Disability benefit level	PIA, earnings, coverage and earnings history of both worker and dependent, year
Spouse benefit	Benefit history of head of family, covered earnings, age, program rules
Children's benefits	Parent's eligibility, age, school status
Survivors Insurance (SI) receipt	Age, coverage, covered earnings of deceased parent or spouse, earnings, timing of parent/spouse's death
Survivors Insurance benefit level	PIA, covered earnings of deceased parent or spouse and dependent, earnings, timing of receipt, year
Family income	Sum of income of all members
(6) <i>Family assets and debts</i>	
Ownership	Lagged ownership, age, race, education, family income, number of children, marital status
Market value	Lagged value, age, race, education, family income, number of children, marital status
Asset transfers	
Divorce	Rule-based: split equally (under development)
Bequests and inheritances	Rule-based (under development)
(7) <i>Taxes on income or assets</i>	
FICA	Covered earnings, year
Federal income tax	Family earnings, family asset income, tax rate
State income tax	Family federal income tax
Property	Family federal income tax

TABLE 5-1 (continued)

DETERMINANTS OF MAJOR EVENTS SIMULATED BY CORSIM

Event or Characteristic	Variables Used to Determine Event ^a
(8) <i>Family consumption and savings</i>	
Consumption expenditures	Family income, number of adults, number of children, age, education, race, weeks worked, marital status, home ownership
Savings	Difference between total family income (after tax) and total family expenditures
(9) <i>Health risk factors</i>	
Cigarette smoking	Age, race, sex, education, family income
Alcohol consumption	Sex, alcohol consumption lagged, marital status, race, smoker, family income, weeks worked, sex
Sugar consumption	Age, race, sex, family income, education, marital status
Diabetes	Age, race, sex, education, family income
(10) <i>Dental insurance and care</i>	
Dental insurance coverage	Age, race, sex, education, year
Dental condition/health	Demographic information, lagged dental conditions, lagged visit to dentist (no feedback to other modules)
Dental services and expenditures	Demographic information, dental condition, lagged visit to dentist, price of services (no feedback to other modules)

^a Data used for each variable are described in Annex 5-1.

^b Events/characteristics are grouped into categories by the author. Categories and order of presentation in this table do not indicated logical relationships in the model or solution order. Solution order is shown in Table 5-2

^c People leaving home for reasons other than marriage, birth of a child, divorce, or death.

Sources: Caldwell, et al. (1993), Caldwell, et al. (1997), unpublished information from Melissa Favreault.

TABLE 5-2**ORDER OF MODULE SOLUTION IN CORSIM**

Beginning of Person Loop

Decide Mortality

Fertility

Aging (only for people not born in current year)

Education

Weeks Worked and Earnings

Sugar (not always activated)

Diabetes (not always activated)

Smoking (not always activated)

Alcohol Consumption (not always activated)

Voting

Disability

Eligibility for Marriage (adds people to the marriage market)

Decide Leaving Home for young adults

Decide who Divorces

End of Person Loop

Immigration

Marriage Mate Matching

Perform Dismemberment for those who died

Perform Divorce for those who were selected

Separate young adults from their families if selected

Wealth Module

AFDC Module (not always activated)

Social Security Module (not always activated)

Beginning of Family Loop

Dental Insurance

State to state migration

Taxes

Expenditures and Savings

End of Family Loop

Most discrete events in CORSIM (e.g. death, birth, marriage) are simulated using a *Monte Carlo* technique.¹⁰ That is, the equation for the event determines a probability that the event will occur for each individual. To determine if the event is assigned to occur, the probability is compared to a random number. Many continuous outcome modules (e.g. wages) are also simulated stochastically. The equation for the outcome produces an expected value for the continuous variable, and a random term distributes the simulated outcome for each individual around the expected value.

CORSIM uses three methods to control unwanted variance that can arise from Monte Carlo simulation: (1) selective sampling, (2) sidewalk method, (3) common random number stream. In *selective sampling* the sample distribution of “random numbers” is contrived to take the desired form by dividing the range of possible outcomes of the probability distribution function into equiprobable intervals, and either sampling from each interval, or selecting the mid-point of each interval.¹¹ In the *sidewalk method* the model keeps a running total of each person’s probability during the course of a simulation run. The person whose probability causes the running total to cross an integer value is designated to have the event occur. Each equation simulated using the sidewalk method is guaranteed to have an actual number of events within one of the expected value. The *common random number stream method* involves the use of the same sequence or stream of random numbers for each source of variation within a model across successive runs of the model. That is, each unit in a simulation receives the same numbers in the same order for each run. If this method is in use, differences between runs would not be produced by differences in random numbers.

Two existing modules -- the wealth module and the social security module -- and a planned module depicting private pensions are of particular relevance for retirement policy.

The Family Wealth Module was developed in 1994. The module was extensively revised under contract to the American Association of Retired Persons (AARP) in 1995, 1996, and 1997.

The Family Wealth Module depicts the ownership and value of each of 11 types of assets and three types of debt. Seven financial assets are represented:

- stocks and mutual funds
- bonds
- checking and savings accounts
- Individual Retirement Accounts (IRAs) and Keogh Accounts
- certificates of deposit and money market accounts
- life insurance assets
- other financial assets.

¹⁰ Monte Carlo simulation is discussed in Chapter 2, p. 2-5.

¹¹ Selective sampling is discussed in Pidd (1988).

Four real assets are represented:

- the primary residence
- other real estate
- vehicles
- business assets.

Three liabilities are represented:

- mortgages on the primary residence
- other mortgages
- all other debts.

Initial values for each type of wealth or liability in 1962 are assigned to each household using the 1962 Survey of the Financial Characteristics of Consumers. Wealth is updated every three years (1965, 1968, 1971, 1974, 1977, 1980, 1983, 1986, 1989, 1992, 1995) for each family using equations estimated from the 1983-86 panel of the Survey of Consumer Finances. Wealth is estimated for each family in two stages. In the first stage, the probability of owning each type of asset or debt is estimated and ownership is assigned by comparing the predicted probability with a random number. In the second stage, for each type of asset for which ownership was assigned, an equation predicts the market value. Variables used in the equations for probability of ownership and value are shown in Table 5-1.

These initial predictions of ownership and value are then aligned to correspond to historical data. First, the initial family estimates are aligned to estimates for demographic groups derived from micro data. For high wealth holders, initial values are aligned using estate tax data.¹² For all other wealth holders, values are aligned using cross-sectional estimates of asset holdings from the Surveys of Consumer Finances (SCF)¹³. Second, the value of wealth holdings for all families are aligned to correspond with historical aggregate wealth estimates from the Federal Reserve Board Flow of Funds Accounts.

For projection years after 1995, wealth ownership and value are assigned to each family using the equations based on the 1983-86 SCF data and alignment data extrapolated from the 1962-1992 time series of alignment data.

¹² Estate tax data were tabulated for the years 1962, 1969, 1972, 1976, 1982, 1986, 1989, 1992 for each of 80 demographic groups to estimate the proportion of persons in each group holding each of eight types of assets and the mean value of asset holdings for each group. Estimates for the missing years were interpolated. The population of top wealth holders was estimated using mortality multipliers derived from a proportional hazards model of mortality using the National Longitudinal Mortality Study. Family wealth was inferred from the estate tax based estimates of individual wealth holdings.

¹³ SCF data for 1962, 1970, 1983, 1986, 1989, and 1992 were tabulated for 40 demographic groups to estimate the probability of holding each asset type and the mean value of each type. Data were interpolated for the update years and demographic groups for which there were no survey data available.

The Social Security Module was developed under contract to the American Association of Retired Persons (AARP). A preliminary version of the social security module was completed in December 1995. Development continued in 1996 and 1997.

The Social Security Module depicts for each individual OASDI coverage, earnings histories, contributions, disability benefit receipt for covered workers and dependents, Average Indexed Monthly Earnings (AIME) and Primary Insurance Amount (PIA), disability benefit amounts, retirement benefit receipt for workers and dependents and amounts, survivors benefit receipt and amounts, and taxation of benefits. Based on the simulated histories of contributions and benefits, the model can produce tabulations of cumulative (lifetime) contributions, benefits, net benefits, ratio of benefits to contributions, replacement rates for various demographic/socioeconomic groups, and other indicators.

The social security module begins by statically imputing pre-1960 earnings histories, including years of coverage and earnings, for all workers in the 1960 data file. It assigns numbers of years of coverage randomly using historic age-sex group probabilities of full coverage. It assigns benefits for workers who began receipt before 1960 based on age-sex proportions in 1960. All beneficiaries are assigned the mean benefit of their age-sex group. CORSIM uses the 1960 census sample as its base data file. CORSIM does not use any of the available social security earnings records data. Consequently, all of the pre-1960 historic information (as well as all information for the period 1960-1996) must be imputed or simulated, even though historic data exist for the 1937-1977 period.

Annual dynamic simulations of OASDI outcomes begin in 1960. Coverage is assigned randomly to wage-earners using historical proportions, without reference to wage level, number of weeks worked, or industry (industry is not simulated). Quarters of coverage are accumulated based on annual coverage assignment and earnings levels. Annual OASDI contributions are calculated based on earnings and the tax rate in each year. Disability (DI) benefit receipt is predicted based on eligibility (number of covered quarters), using an equation estimated with PSID data, which includes (self-reported) disability status and lagged receipt of disability benefit as explanatory variables (see Table 5-1). DI benefit receipt for dependents is determined based on the relationship to the worker beneficiary. AIME/PIA is calculated for workers with sufficient covered quarters who are slated to receive a benefit. AIME is calculated using the social security indexing formula and the worker's simulated earnings history. PIA is calculated from AIME based on the actual or projected formula. Disability benefit amount is assigned based on the PIA and the family cap rule. Retirement benefit receipt is predicted for workers ages 62-69 using an equation estimated with PSID data for 1986-1991 (see Table 5-1). Explanatory variables in the retirement equation include demographic characteristics, current work hours and wages, and asset income, but do not include variables for prospective social security benefit or pension benefit levels. Retirement benefit amount is then calculated based on the PIA and adjusted for early or late retirement and the family cap. Survivors benefit receipt is based on the time of death of the worker, and amount is determined by the worker's PIA. OASDI trust fund balance is calculated based on total contributions paid in, taxes on benefits, and total benefits paid out. Interest earned on the fund, reimbursements, and administrative expenses are exogenous.

The social security module produces voluminous output, which can be tabulated in many ways, including annual counts of contributors, amount of contributions, counts of beneficiaries of various types in various socioeconomic groups, total benefit payments, average benefits. Longitudinal (life course) results can be tabulated for cohorts, age groups, or other socioeconomic groups (e.g. income before retirement, wealth, marital status over various periods, race, sex).

Because social security (as well as employer pensions) is very sensitive to life course employment and earnings histories, both individually and in the aggregate, the issue of the validity of CORSIM imputations of coverage and earnings data for 1937-1959, and simulations of historic employment and earnings data for 1960-1996, rather than the use of actual data, is important. Social security and retirement income patterns depend on the joint distributions of employment and earnings and many other socioeconomic attributes, such as age, sex, marital status and marital histories, childbearing, and disability, as well as the expected or mean levels of these attributes for various demographic groups.

As of October 1997, simulations and projections of social security aggregate revenue, benefit expenditures, and trust fund balances had not been made available to be compared with historic data and projections made by others. CORSIM simulations of historic per capita benefit levels and numbers of recipients in various socioeconomic categories for the 1960-1995 period differed substantially from reported data.

Proposed Private Pension Module. CORSIM staff have designed a Private Pension module, but it had not been developed as of September 1997.¹⁴ The design of the CORSIM Private Pension module includes functions to estimate pension coverage and benefit entitlement and to estimate the disposition of pension rights in the event of death, divorce, job change, disability, and retirement. The Private Pension module creates a pension record for each CORSIM person. The pension record includes a description of the individual's pension plan(s) characteristics and participation history (contributions and years of service). Each year, the person's pension record is updated based on his/her employment and whether or not he or she dies, divorces, or reaches the terminal age for receiving benefits.

Private pension coverage and benefit entitlement are a direct result of a person's employment. If a person has any weeks of work during the year (determined by the CORSIM weeks module), the Private Pension module first imputes the characteristics of the person's employment, including full-time/part-time status, union membership, and whether the individual is self-employed. The module also determines if the weeks worked in the current year were at the same or a different job from the previous year. If the person has changed jobs, the module determines the disposition of previously accumulated pension benefits, if any. For all persons who had any weeks of work in a given year the module determines pension coverage and participation. For persons who were covered by a pension

¹⁴ This module is to be developed in 1997 and 1998 under contract to the American Association of Retired Persons (AARP).

last year but have no weeks of work in the current year the module determines the disposition of any accumulated pension assets.

The Private Pension Module includes seven sub-modules:

- Pension Coverage - updates a person's pension record based on employment status for the current year;
- Retirement - determines pension retirement benefit amount and payment type;
- Disability - determines disability benefits amount;
- Death - determines death benefits and survivorship benefits;
- Pension Benefits - determines benefit payments eligibility for disabled and dependent benefit recipients;
- Divorce - determines division of pension rights;
- Termination - determines disposition of pension accruals after termination of employment (other than retirement).

The structure of the proposed Private Pension Module is shown in Annex 5-3.

The Pension Coverage Sub-Module performs three functions: (1) determine if a person has pension coverage in a given year; (2) assign/determine pension plan characteristics for new plans; (3) update pension record for individuals continuing pension participation.

If a person is determined to be covered by a pension plan, the module will assign a type of plan (defined benefit, defined contribution, or both) and then assign plan characteristics as follows:¹⁵

Defined benefit

Contributory/non-contributory

If contributory, assign salary percentage

Salary base for determination of pension benefit

Average of last *x* years of employment

Average of highest *x* years of employment

Career average

Integration with Social Security (Y/N)

Replacement rate (percentage of salary base pension will 'replace')

Defined contribution

Maximum and minimum contribution percentages

Employer match/no employer match

If employer match, assign maximum match percentage

Rate of return

¹⁵ CORSIM's pension module design uses the DYNASIM approach of assigning parameters of a prototypical pension plan to each participant, rather than the PRISM approach of assigning participants to actual plan sponsors and using the provisions of those sponsors' actual pension plans.

All plans

- Vesting provisions
 - Cliff (vested after x years of service)
 - Full and immediate
 - Graded ($x\%$ at year y , $x\%$ at year $y+t$, etc.)
- Spousal rights on death of pensioner
 - spouse pension of $x\%$ of employee's pension
 - continuation of employee's pension for x months
 - joint survivorship pension
- Dependent benefits on death of pensioner
- Death benefits (lump sum payment on death)
- Disability benefits
 - qualifications
 - benefits
- Normal retirement age
- Early Retirement provisions.

Each person that has a pension will have a pension record that includes (for each pension) the characteristics of the pension and the individual's qualified years of service under the plan. Additionally, for contributory plans, the record will include the employee's yearly contribution and employer match contribution, if applicable. Persons with defined contribution plans will also have an entry for the annual rate of return. These entries will be updated yearly.

The Retirement sub-module determines the pension benefits received by persons upon their retirement. It is assumed that the defined benefit plans will pay a monthly benefit, calculated according to the characteristics recorded in the pension record, and that defined contribution plans will pay out a lump-sum which is rolled over (tax free) to an annuity.

For defined benefit plans, the module determines if the person meets the pension plan requirements for "normal" retirement; either by reaching retirement age (usually 65) or reaching a minimum number of years of service (usually 20). If not, the plan calculates the early retirement monthly benefit. Then, the module checks for a defined contribution component and calculates the lump-sum benefit if applicable. For defined contribution (only) plans, the module calculates the person's lump-sum distribution.

The Disability sub-module calculates monthly benefits from private pensions for individuals that become permanently disabled. The Death sub-module assigns the disposition of pension benefits or rights upon the death of the covered person.

The Pension Benefits Sub-Module performs the following functions: (1) adds retirement (for defined benefit pensioners), disability, spouse, and dependent benefit payments to income; (2) moves disabled persons to retirement benefits when they reach the normal retirement age for their plans; and (3) ends dependent pensions when the beneficiary reaches 18 years of age.

The Divorce Sub-Module assigns the disposition of pension rights upon divorce. The Termination Sub-Module determines the disposition of pensions when an individual changes jobs or stops working for a reason other than retirement.

As of October 1997, CORSIM staff had identified several potential data bases to develop the parameters of the pension model, but had not determined which ones to use. Under consideration were the Survey of Consumer Finances, the Current Population Survey Employee Benefits Supplements, the survey of Asset and Health Dynamics Among the Oldest Old (AHEAD), the Health and Retirement Survey (HRS), the Retirement History Survey, and the PSID.

Controls and Alignment

CORSIM has an extensive set of alignment procedures for most modules. Because CORSIM begins with a 1960 data file and simulates the period 1961-1996, extensive historical data are available to compare and align the simulated data. Because CORSIM's projections for future years are based largely on simulated historical data, it is important that the simulated historical data be aligned to actual data. Most module outcomes are aligned in two stages.

First, individual micro outcomes (usually probabilities) are aggregated or tabulated for demographic groups for which independent historical data are available. Group alignment factors are calculated by comparing the group mean of the initial simulated outcomes with the historical group data. Individual outcomes or probabilities are then adjusted so that the group mean after alignment corresponds to the historical data. The adjustments are usually multiplicative. This alignment procedure requires a second pass through the data in the file. In the first pass the initial outcomes or probabilities are estimated, which are used to calculate the alignment factors. In the second pass the initial micro outcomes are adjusted to align to the historic data. Because probabilities must be constrained to be between zero and one, the adjustment may take the form of an additive adjustment to the logit of the probabilities, rather than a multiplicative adjustment to the probabilities themselves, which could make some probabilities greater than one.¹⁶ This non-linear alignment procedure may require more than one iteration.

In the second stage, the outcomes aligned to group data are aggregated and compared to historic aggregate data. The individual outcomes may then be aligned so that in the aggregate they match the historic aggregate data.

¹⁶ Most probabilities in CORSIM are estimated using a logistic specification: $p = 1/(1+e^{-Bx})$. Parameters are estimated by transforming to the logit: $Bx = \ln(p/(1-p))$. The micro equations are then aligned by comparing the logit calculated using the historic group data with the logit calculated using the mean of the initial simulated probabilities. Adding the difference between these two "group-level" logits to the logit equation for each individual will not, in general, produce an exact alignment on the first pass, but ordinarily converges after several iterations. CORSIM documentation refers to the group mean of the initial simulated probabilities as the *micro rate* or *mrate* and the historical group average rate as *hist*. The adjustment constant to be added to the Bx is $gp = \ln(hist/(1-hist)) - \ln(mrate/(1-mrate))$.

This alignment procedure is designed so that the historic simulations track historic data, by aligning to historic group data, while the “distribution” of individual outcomes over individuals in the group reflects individual-level variables.

CORSIM uses the time series of historic alignment factors to develop alignment factors for the projections of future years. In addition, projections of future years can be aligned to other projections of aggregate outcomes, such as aggregate fertility rates, mortality rates, rates of average wage growth, etc. Simulations are often aligned to projections done by the Office of the Actuary of the Social Security Administration.

Databases

CORSIM operates on the 1960 U.S. Census Public Use Microdata Sample (PUMS). The full PUMS is a one-in-1000 sample of the U.S. Census and includes about 180,000 person records.¹⁷ Because it is a random sample of the census, each of the person records in the PUMS has the same weight, 1,000, meaning that each person record in the PUMS represents 1,000 persons in the U.S. population. Often, CORSIM is run on smaller samples of the PUMS, such as 20,000 or 50,000.

Data from nine large national microdata files were used to estimate the equations of the model, and data from other files, including survey data and program administrative data, were used to develop rule-based algorithms and to validate the model. The data used to develop and operate the various modules of the model are listed in the Summary Description Table in Annex 5-1. Databases used to develop demographic modules include the National Longitudinal Survey (NLS) (1973-87), National Longitudinal Survey - Youth (1969-87), National Longitudinal Mortality Survey (1980-89), Panel Study of Income Dynamics (PSID) (1968-88), High School and Beyond Survey (HSB) (1980-86). Work and earnings modules were estimated using the PSID (1972-88). Asset accumulation was estimated and aligned using the Surveys of Consumer Finance of 1962, 1966, 1971, 1975, 1983, 1986, 1989, 1992, as well as IRS estate tax data and the Federal Reserve Board Flow of Funds Accounts. Health risk factors were estimated using National Health Interview Surveys of 1970 and 1984 and the NLS 1986-87. In addition, census data of 1960, 1970, 1980, 1990 were used.¹⁸

CORSIM’s use of the 1960 Census PUMS as its base data file is an issue. A large data file that is several decades old was chosen to serve as the base file to permit CORSIM to simulate the evolution of the population over several decades for which historic data are available, so that the operation of the model could be assessed by comparing its simulations to actual data and to use the model to investigate and understand actual historic outcomes. This reflects CORSIM’s origins and primary role as a tool for academic research. The CORSIM developers argue that simulating the

¹⁷ The population of the United States in 1960 was 179,323,175.

¹⁸ CORSIM staff are considering several databases for development of the private pension module, including the Survey of Consumer Finances, Current Population Survey Employment Benefits Supplements, AHEAD, Health and Retirement Survey, Retirement History Survey, and PSID.

model for a lengthy historic period permits them to understand and validate the socioeconomic processes depicted in the model and develop alignment factors that will improve projections of future years. However, CORSIM simulations of many processes do not track historical data very accurately, even after the first stage alignments (described above). Without realigning or rebasing the data for a recent historic year, projections of future years may begin from a base that already is subject to errors accumulated over a 35 year simulation period. Even if many group and aggregate outcomes can be aligned to recent data exactly, there is no way to assure that the joint distributions based on the 1960 data remain accurate after 35 years.

Documentation

Although extensive and informative, CORSIM's documentation is incomplete and inadequate. Available documentation includes a lengthy 1993 research documentation for CORSIM 2.0¹⁹ and a comprehensive 1997 technical and programmer's guide for CORSIM 3.0.²⁰ The research documentation contains interesting and informative background discussion of microsimulation modeling and the issues and problems involved, and background discussion of some of the modules. The discussion of many modules is incomplete and out of date. It provides almost no empirical estimation results, so it is impossible for the reader to evaluate the validity of the model equations.

Computer Hardware, Software and Portability

A key objective of CORSIM was portability across computational platforms. That dictated the choice of ANSIC language and the requirement that programmers write portable code. CORSIM runs under OS/2, UNIX, or Linux 1.3. A sophisticated user-friendly interface has not been developed. CORSIM is highly modular and designed so that existing modules can be modified and new modules can be added relatively easily.

CORSIM 3.0 currently is run at Cornell University on a dual processor Pentium Pro, 200 MHZ each, with 4 GB hard drive and 256 MB RAM, as well as a main frame (IBM 3090/600). On the dual Pentium Pro a CORSIM run of 130 years (1960-2090), including the social security module, excluding the health risk, dental, voting and consumer expenditures modules, on a 50,000 person sample takes about fourteen hours. A run of the wealth module on a sample of 180,000 for 30 years takes about four hours.

¹⁹ Caldwell, Steven B., *CORSIM 2.0 Model Documentation Version 5*, Cornell University, Ithaca, NY, October 1, 1993.

²⁰ Caldwell, Steven B. *et al.*, *CORSIM 3.0 User and Technical Documentation*, Cornell University, Ithaca, NY, March 1997. This document is available on the CORSIM web page at <http://mistic.soc.cornell.edu/docs/techdoc/index.html> (index) and http://mistic.soc.cornell.edu/docs/techdoc/new_tech.html.

IV. APPLICATION TO IMPORTANT RETIREMENT POLICY ISSUES

The policy matrix tables in Annex 5-2 indicate the types of issues which CORSIM is suitable to address. There is a table for each of six major areas where changes in retirement income policy could have important implications. The rows, which are the same for each table, list several important areas where policy changes could occur. The columns, which differ in each table, show aspects which are relevant or of concern in each of the major areas. Entries in each table show areas where CORSIM is suitable for analysis of the impacts of the types of policy changes indicated in the rows on the aspects of the pension system indicated in the columns.

CORSIM models individual person and family behavior. It can be used to simulate individual work histories, social security contributions, social security benefits receipt and amounts, and family wealth holdings. CORSIM records extensive kinship links, including links among divorced family members. It can analyze some policy issues for which information on personal histories is required. The current version of CORSIM does not model employer pensions, so it cannot address policy issues pertaining to private or government employee pensions.

Examples of issues suitable for analysis with CORSIM include social security policy issues such as earnings sharing, changes in benefit formula and indexing, and adoption of individual accounts; effects of marriage and fertility changes on AFDC; effects of changing schooling on weeks worked and wages. CORSIM does not simulate employer or plan sponsor behavior, so it does not have the capability to analyze the effects of policy measures on employers or on their offering of pensions or other benefit plans. It does not depict the behavior of producers or industries, so it cannot be used to analyze industry outcomes. Other than very simple labor input variables, individual weeks worked and wages, CORSIM does not depict any aggregate economic behavior, so it cannot analyze effects on the aggregate economy, such as saving, investment, GDP growth, or interest rates.²¹

Benchmarks

Appendix C describes seven illustrative policy issues that can be used as benchmarks to assess and compare the suitability of various models for analysis of retirement income policy in various areas. This section reviews the suitability of CORSIM for analysis of each of these illustrative benchmark policy issues.

1. **Effects of increase in Social Security Normal Retirement Age on:**

OASI revenues, benefit payments, trust fund balances -- CORSIM has the formal mechanics to analyze the statutory effects of change in social security normal retirement age (NRA). Change in

²¹ The two labor input related variables --weeks worked and wages -- are modeled as functions of socioeconomic characteristics of the person, e.g. age, race, sex, education, marital status, etc. No labor demand-related factors are included, and no aggregate economic indicators are currently activated.

eligibility for OAI full retirement benefit and change in early retirement reduction can be programmed and will change the retirement benefit amount. However, CORSIM does not have a behavioral retirement model. None of the variables affected by the change in normal retirement age are represented in the CORSIM retirement benefit receipt timing equation. *Ad hoc* assumptions about potential changes in retirement behavior could be made. In concept, CORSIM can be used to calculate effects on OASI revenues, benefit payments, and trust fund balances, but the trust fund model is rudimentary. It can calculate payroll tax liabilities and social security retirement benefit payments for each person, which can be aggregated.

DI benefit payments and trust fund balances -- The CORSIM DI module simulates increased eligibility for DI benefits as the NRA is increased under the 1983 law. (The methodology is not documented.) Accurate analysis may require additional research. Currently, all DI beneficiaries convert to OAI at age 65. The DI module was not estimated for persons age 65 and 66, since there are no historic data for DI receipt at ages 65 and 66.

Employer pension accruals and benefit costs -- CORSIM does not model employer pension plans or employer behavior. Pension accruals and benefit costs are not depicted.

Social security retirement replacement rates and total retirement income replacement rates -- CORSIM's longitudinal earnings records and social security model provide the capability to calculate social security replacement rates, for various assumptions about retirement behavior. CORSIM does not currently depict employer pensions. The family asset module does not depict effects of changes in retirement behavior or expectations on assets. Consequently, CORSIM could not analyze effects on total retirement income.

2. Means testing of Social Security benefits.

In concept, with some reprogramming, CORSIM could be used to estimate the effects of means testing of social security benefits on total benefit payments. It depicts earnings and asset income, but currently it does not depict pension income. It currently depicts the effects of the social security earnings test on individual benefits and means testing of SSI. CORSIM does not have behavioral models of DI or OAI benefit receipt. *Ad hoc* assumptions about benefit receipt would be required. Effects on income replacement rates could be simulated. CORSIM does not depict employer costs, so the effects on employer costs could not be simulated.

3. Mandatory minimum employer pension.

The 1997 version of CORSIM could not be used to estimate effects on workers or employers of a mandatory employer pension. The proposed private pension module may have the capability to simulate effects of a mandatory minimum pension on workers. CORSIM could not simulate the effects on employers or the potential changes in other pensions, wages, or employment. CORSIM does not depict employer pension accruals.

4. Expansion of individual retirement account eligibility.

CORSIM could not be used to estimate effects of expansion in IRA eligibility. CORSIM does not have an individual retirement account model. Individual retirement account assets in the family wealth module are not modeled with reference to IRA rules. Because CORSIM does not depict saving behavior or macroeconomic effects, it could not estimate any policy effects on total retirement

savings, personal savings, national savings, capital formation, or GDP. It does not depict federal expenditures, deficits, or debt.

5. Effects of value added tax on pension contributions and accruals.

CORSIM cannot address issues concerning the effects of taxes on pension offerings, contributions, or accruals. The model does not depict the effects of taxes on individual or firm behavior. It does not depict the behavior of corporations or other employers, so it does not depict pension plan sponsors.

6. Effects of construction industry benefit accrual rates on funding.

CORSIM is not suitable to analyze issues concerning the condition or behavior of employer pension funds. CORSIM does not model employers, industries, sponsors, or pension funds.

7. Effects of alternative macroeconomic scenarios on social security and employer pensions.

CORSIM in general is not appropriate for analysis of effects of alternative macroeconomic scenarios on other aggregates, such as social security fund balances. Aggregate economic variables do not enter any individual equations. Individual labor market outcomes in each period, such as weeks worked and wages, can be controlled to align to external group and aggregative data. Projections could, in concept, be controlled to align to alternative macroeconomic scenarios. CORSIM could in this way be used to simulate longitudinal earnings and employment histories under alternative macroeconomic scenarios. These longitudinal earnings and employment histories are inputs into the determination of social security receipt and benefit levels. In this way CORSIM can be used to depict some of the effects of alternative macroeconomic scenarios on various socioeconomic or demographic groups. CORSIM does not depict employer pension fund balances or accruals. The 1997 version of CORSIM does not depict individual participation in or receipt of employer pensions.

V. ACCESSIBILITY AND EASE OF USE

CORSIM is a proprietary model owned by Cornell University. CORSIM was developed to facilitate research in microsimulation modeling, and the CORSIM project wants to encourage use of the model. Model code is available to users at reasonable cost and to collaborators at no cost. The code is designed to be portable. The technical user's guide provides instructions in the implementation, use, and modification of the model. The code is written in ANSI C programming language, and a technical user planning to adapt the model should be familiar with C. The model does not have a user-friendly interface, but the code and technical documentation are reasonably clear and accessible to the technical user familiar with C. It is likely that any outside individual or organization acquiring the model would require assistance initially from CORSIM programming staff, which could be provided at reasonable cost.

The available documentation of the model for the research or policy analysis user is out of date and unsatisfactory (see Section III). Considerable effort would be required for an analyst unfamiliar with the model to understand its current features and properties and to assess the validity of its modules.

CORSIM may provide a platform with potential to add modules of interest to the Society of Actuaries. It was designed to permit researchers to add or modify modules. It is highly modular and portable. The CORSIM team at Cornell are currently developing a private pension module. The private pension module is still in the design stage, and its developers may be open to input and assistance from outside groups.

CORSIM does not provide any representation of firms, employers, government entities, industries, or other organizations. Consequently, development of modules depicting plan sponsor or employer behavior would require a major development effort. It has no regional attributes.

The CORSIM structure and computer code have been successfully adopted and adapted by the Office of the Superintendent of Financial Institutions (OSFI) of the Government of Canada to serve as the basis for the development of a model of the Canadian population and Canada Pension Plan (which corresponds to the U.S. social security system).²² CORSIM staff and OSFI staff continue to work together on development and updating of both models.

Contracting with the CORSIM project to conduct specific studies, add modules, or provide the model and assistance in its use would be feasible. CORSIM staff appear to be interested in collaboration with outside groups.

The fact that CORSIM has been developed and is maintained at an academic institution may be associated with certain features that should be considered. It has been primarily a research tool. This orientation accounts for the use of a 1960 data file to serve as the base population sample to be aged. Valuable research results and insights can be developed from investigation of simulations of past periods, even if those simulations do not replicate closely historical data. However, greater verisimilitude may be a more important property in a policy analysis environment to lend the analysis credibility. CORSIM staff have been composed of a full-time faculty member and graduate students (primarily in sociology) at Cornell. CORSIM has benefitted from the ability to draw on resources and personnel at a major research university, and research conducted by faculty and graduate students has contributed to its development. However, CORSIM does not have a permanent staff of experienced professionals. It should be recognized that academic personnel may have multiple commitments and may not always be available full time to an outside organization.

VI. CRITIQUE

CORSIM is a large scale microsimulation model which simulates many demographic and social events, giving it flexibility and breadth. The computer code is portable and modular. The technical user/programmer documentation appears to be fairly clear and complete, although in 1997 it did not include the current wealth nor social security modules. The documentation for the researcher/analyst is incomplete and out-of date.

²² This model is described in Chapter 6.

CORSIM includes a comprehensive social security module which depicts in detail the provisions of the system and has extensive report and tabulation capabilities.

CORSIM has relatively little economic behavioral content. With the exception of the weeks worked and wages modules developed in 1988, neither economists nor actuaries have worked on CORSIM. Most of the equations for attributes and behaviors in CORSIM are descriptive. In general, the equations do not appear to be based on well articulated theories or models of economic behavior.²³

From the perspective of retirement policy, CORSIM has several weaknesses. The labor market module is weak. The only two aspects of labor market behavior that are depicted are weeks worked and wages, and these are modeled with reference to demographic attributes only (see Table 5-1).²⁴ There is no depiction of labor force participation, unemployment, or hours. There is no identification of industry affiliation or size of firm of workers, two attributes that are important for pension analysis. Employment class (self-employed), occupation, labor union affiliation are not depicted.

The social security model assigns coverage randomly (aligning to historical percentages). The model uses no historical earnings records, but rather imputes earnings before 1960 to each worker, and simulates earnings after 1960. This is an area where data are available, and where accurate imputation would seem to be both important and difficult.

CORSIM has no behavioral retirement function. In the current version, there is no feedback from changes in social security or other aspects of retirement income or retirement saving to affect labor market behavior or retirement.²⁵

²³ Of the major microsimulation models, DYNASIM probably incorporates the most economic research, although much of this research may now be out of date. In fact, all three microsimulation models reviewed in this report have many components that are *ad hoc*, in the sense that they have equations or transition probabilities that describe or replicate data in a particular data base or observed trends over a particular period, without providing a more general, underlying explanatory model. If the underlying determinants change in the future, the descriptive equations or probabilities may no longer be accurate.

²⁴ The aggregate unemployment rate was included in the estimation equation for annual weeks worked, but it is not activated in the current version of the model.

²⁵ A mechanical or rule-based, rather than behavioral, approach to modeling social security is reflected in the description of the CORSIM social security module. The authors write:

We have attempted to construct the OASDI module in such a way that any step upon which other steps depend precedes the dependent step or steps. For example, *assignment of retirement benefit timing must always precede assignment of retirement benefit level*, given that there are adjustments to benefit amounts mandated for cases of early or delayed retirement. As essentially all of the steps in CORSIM's OASDI module depend upon marital, child rearing, and labor force trajectories, *the module is invoked only after outcomes for all of CORSIM's regular processes (deaths, births, marital status changes, changes in work effort, and so on) have been determined for the current year.* (Caldwell and Favreault, 1995) (Italics added by the author.)

Similarly, CORSIM has an elaborate family wealth imputation module, but it has minimal economic behavioral content and does not provide links to social security or other aspects of retirement income or retirement savings policy or institutions.

A crucial aspect of CORSIM is its reliance on the 1960 census sample for its basic data file. From the perspective of its developer, this is a strength and provides a framework for valuable research into both socioeconomic patterns and microsimulation modeling methodology. It provides opportunities to investigate the accuracy of various modules, theories, and behavioral representations, and to improve the model's alignment to historical data. This view holds that the opportunity to examine the behavior of the model over a long historical period will improve its validity for projecting future periods. However, the ability of the model, even with extensive alignment, to provide simulated historical data that are accurate and capture the many changes in family structure, socioeconomic patterns, and joint distributions that have occurred over the past 35 years is a concern. In particular, the use of randomly assigned social security coverage and simulated earnings histories for the 1960-1996 period may be a particular concern for retirement income policy. (It must be recognized, however, that the most recent available social security earnings histories data are from the 1978 CPS-SSA match, and are 19 years old in 1997, so earnings histories must be simulated for the 1979-1996 period in any case.)

Despite alignment to historic group data, the reported simulations of social security outcomes and family wealth appear to vary significantly from historic aggregate data. This attests to the perhaps inherent difficulties of microsimulation modeling. It also reinforces the concern about using simulated historic data as the base for projections of the future.

Of the three major U.S. microsimulation models reviewed in this report, CORSIM has the most recently developed code and simulation system. It does not yet have a private pension module, although one is in the design stage. The CORSIM private pension model design is similar to that of DYNASIM, in that it assigns workers to simplified, prototypical pension plans, rather than to actual plan offerings of actual plan sponsors, as does PRISM. Of the three microsimulation models, CORSIM has the least economic behavioral content. It also provides the least information relevant to retirement policy, such as industry, hours worked, firm size, and labor union affiliation. CORSIM has been designed to be portable and to be adaptable and expandable and to investigate the feasibility of conducting microsimulation research and developing microsimulation models at reasonable cost.

5-26

ANNEX 5-1

CORSIM 3.0

SUMMARY DESCRIPTION TABLE

ANNEX 5-1

CORSIM 3.0

Summary Description

Subject: U.S. person and family socioeconomic conditions and behavior.

General Objectives of Model

- **Basic social science research:** investigate the processes underlying inter-generational and intra-generational socioeconomic mobility; life course patterns; dental care utilization and outcomes; demographic and spatial dynamics; income distribution and wealth accumulation; and other topics related to the socioeconomic behavior of persons and families in the U.S. since 1960;
- **Policy analysis:** simulate the consequences of alternative design features of tax, income-conditioned transfer and social insurance programs, and the effects of a range of socioeconomic conditions on program outcomes.

Specific Purposes of Model

- Dynamically age a large sample representing the U.S. population -- the full 1960 U.S. Census Public Use Microdata Sample (PUMS) of 180,000 persons -- over the 1960-1996 historical period and 1996-2030 future period (1996-2090 for Social Security analyses);
- Create a series of synthetic annual cross-section microdata files with detailed socioeconomic attributes for each person and family in a sample representing the U.S. population;
- Create longitudinal microdata files with detailed socioeconomic histories for each person and family in a sample representing the U.S. population;
- Synthesize the empirical strengths of numerous, diverse data files of varying types (e.g., longitudinal microdata, aggregate totals, cross-section microdata, vital statistics, administrative statistics) in creating simulated cross-section and longitudinal files;
- Carry out “what if” experiments covering both historical and future periods to assess the cumulative impacts of specific behaviors and program parameters;
- Analyze public policies in which cross-sectional and life history information about families and individuals plays an important role.

Period of historical analysis: 1935-1960, and selected data for 1962, 1966, 1969, 1970, 1980, 1992, 1969-87, 1980-89, 1968-88, 1986-91

Forecast/simulation horizon: 1960-2030 (2090 for OASDI analyses)

Frequency: Annual

Base year: 1960

Simulation technique: Dynamic microsimulation with iterative alignment

Solution algorithms and structure: Sequential/recursive, single equations, transition matrices

Unit(s) of analysis: Persons and (nuclear) families

Cell Structure: Individual person and family records from Census PUMS, weighted to represent U.S. population

Databases**Base Year Database:**

1960 U.S. Census Public Use Microdata Sample (PUMS),
180,000 person records.

Processes/Modules**Demographic processes**

- fertility
National Longitudinal Survey-Youth, 1969-87 (28 groups);
Vital Statistics (26 groups); OASDI Trustees Report total
fertility rate (TFR)
- immigration
U.S. Immigration statistics
- mortality
National Longitudinal Mortality Survey, 1980-89 (51 groups);
Vital Statistics (88 groups); OASDI Trustees Report age-
sex adjusted death rate
- first marriage
National Longitudinal Survey, 1973-87 (20 groups); Vital
Statistics (16 groups)
- assortative mating
Decennial Censuses: 1960, 1970, 1980
- divorce
Panel Study of Income Dynamics (PSID), 1968-88 (4
groups); Vital Statistics (14 groups)
- remarriage
Panel Study of Income Dynamics, 1968-88 (7 groups); Vital
Statistics (26 groups)
- custody of children
Various
- leaving family of origin
High School and Beyond (HSB) Survey, 1980-86
- institutionalization
Various survey and census data

Family/kinship linkages

1960 Census; thereafter determined by model evolution

- among parents and children
- among spouses (& ex-spouses)
- among siblings (& half- and step-siblings)

Geographic mobility

Decennial Censuses: 1960, 1970, 1980, 1990

- migration out of state of residence
- migration into new state of residence

Schooling

High School and Beyond (HSB) Survey, 1980-86; National
Longitudinal Survey, 1979-87; Dept. of Education data

- enroll in school
- progress through ninth grade
- continue to twelfth grade
- graduate from high school
- enroll in college
- continue to second year of college
- continue to third year of college
- continue to fourth year of college
- graduate from college
- enroll in post-graduate education

- continue to second year of post-graduate
- continue to third year of post-graduate
- complete more than three years post-graduate

Work and earnings Panel Study of Income Dynamics, 1972-87 (174 groups for work status (0/FT/PT), 58 groups for weeks worked, 116 groups for earnings); Decennial Censuses: 1960, 1970, 1980, 1990; CPS 1971, 1981, 1991 (35 groups for weeks worked, 70 groups for earnings); OASDI Trustees Report intermediate coverage rate (2 groups); National Income and Product Accounts; OASDI Trustees Report intermediate assumption on real wage differential

- annual weeks worked
- annual earnings

Asset accumulation 1983-1986 Surveys of Consumer Finance; IRS Estate Tax Files for 1962, 1966, 1969, 1972, 1979, 1982, 1986, 1989, 1992 (80 groups); National Longitudinal Mortality Survey; Surveys of Consumer Finance for 1962, 1966, 1971, 1975, 1983, 1986, 1989, 1992 (70 groups); Survey of the Financial Characteristics of Consumers (1962); National Flow of Funds Accounts

- home (primary residence) ownership, home value
- other real estate ownership, other real estate value
- business asset ownership, business asset value
- vehicle ownership, vehicle value
- stocks and mutual funds ownership, stocks and mutual funds value
- bonds ownership, bonds value
- checking/savings account ownership, checking/savings value
- IRA/Keogh account ownership, IRA/Keogh value
- CD/money market account ownership, CD/money market value
- life insurance assets ownership, life insurance assets value
- other financial assets ownership, other financial assets value
- home mortgages ownership, home mortgages value
- other mortgages ownership, other mortgages value
- all other debts ownership, all other debts value

Asset transfers and taxes

- asset transfers at death IRS Estate Tax Files
- asset transfers at divorce Various
- estate taxes Tax laws

Disability

- Self-reported disability PSID 1986-1991
- Receipt of SSDI (timing) PSID 1986-1991
- Receipt of SSI-Disabled (timing) PSID 1986-1991

Risk factors

- smoking National Health Interview Survey (NHIS), 1970, 1984
- alcohol National Longitudinal Survey (NLS), 1986-87
- sugar National Health and Nutrition Survey (NHANES), 1984
- diabetes National Health Interview Survey (NHIS), 1970, 1984

Social Security (OASDI)

OASDI regulations; aligned with various data; some stochastic functions from PSID, 1986-1991

- OASDI coverage
- OASDI contributions
- Calculation of AIME/PIA
- Disability benefits receipt: worker
- Disability benefits receipt: dependent spouses and children
- Disability benefits: amount
- Retirement benefits receipt: worker
- Retirement benefits receipt: dependent spouses and children
- Retirement benefits receipt: conversion from disability
- Retirement benefits: timing
- Retirement benefits: amount (includes deductions for early retirement and credits for delayed retirement)
- Survivors' benefits receipt: timing
- Survivors' benefits: amount
- Taxation of OASDI benefits since 1984
- Cumulative (i.e., life course) outcomes include: internal rate of return, lifetime transfer/tax, ratio of benefits to contributions, replacement rate (3 versions); proportion of adulthood/retirement in post-OASDI poverty

Other programs

National Longitudinal Survey, Youth (NLS-Y) 1979-1991; PSID, 1986-1991; laws and regulations: administrative data

- Supplemental Security Income (SSI), both aged and disabled programs: receipt (100 groups)
- Supplemental Security Income (SSI), both aged and disabled programs: amount
- Earned Income Tax Credit (EITC): receipt
- Earned Income Tax Credit (EITC): amount
- Aid to Families with Dependent Children (AFDC): receipt (6 groups)
- Aid to Families with Dependent Children (AFDC): amount
- other employment-related transfer programs: receipt/amount
- other welfare-related transfer programs: receipt/amount
- other pension-related transfer programs: receipt/amount

Private pensions (planned for 2001)

Pre-tax family income	By identity
<ul style="list-style-type: none"> ● total family earned income ● total family transfer income ● total family asset income ● total family pre-tax income 	
Taxes	Tax modules based on tax rules.
<ul style="list-style-type: none"> ● payroll taxes ● federal income taxes ● state income taxes ● property taxes 	
Disposable family income	By identity
<ul style="list-style-type: none"> ● post-tax income 	
Consumption and savings	1990 Consumer Expenditure Survey
<ul style="list-style-type: none"> ● consumption ● savings 	
Initial individual/family characteristics	1960 Census, and various other imputations
Employer characteristics	None
Industry characteristics	None
Retirement plan characteristics	OASDI only; private pensions planned in 1997
Macroeconomic data	Macroeconomic controls provided exogenously.
Retirement Behavior	Implied by 'weeks worked' equations only; logistic equation for beginning receipt of OASDI
Health conditions	Dental only; DENTSIM module connects to CORSIM 3.0 and simulates use of detailed dental services and expenditures, and dynamically updates detailed dental status of each tooth, gum and mouth
Health insurance coverage	Dental insurance only

Data Quality

Completeness: For characteristics depicted, data are sufficiently complete. (Employer pensions, health other than dental are not depicted.) Eighteen different national data files used in the estimation of module parameters. The 1960 PUMS data serve as the initial file. Administrative data and laws are used extensively.

Accuracy: Base year file used is 1960 Census sample. Micro outcomes (states, events), aggregated to match historical totals and subtotals, are in most cases (and all important demographic ones) aligned iteratively to historical totals and sub-totals over the 1960-1995 period, or at least as close to 1995 historical values as are available. Accuracy of distributions and covariances for years removed from 1960 is uncertain.

Representative: The initial data base is a 1960 Census sample, and is representative of the U.S. population in that year. Alignment of aggregates from 1960 to 1995 is designed to maintain representativeness, but this is an issue.

Currency: Base year file is 37 year old in 1997. As explained above, outcomes are adjusted to match more recent aggregate data. Currency of results is uncertain.

Applicability to other contexts: CORSIM 3.0 is very comprehensive and can be applied to many different issues and contexts.

Applicability of other private/consulting firm data: Data on employer pension plans could potentially be used to help build the planned CORSIM 3.0 pension module.

Gaps: No earnings history data are used.

Other characteristics, activities, behaviors that are modeled

Economic activity Not modeled. Controls provided exogenously (typically the 1996 OASDI Trustees Report Intermediate Assumptions).

Short-run/cyclical: exogenous

Long-run growth, productivity: exogenous

Inflation: exogenous

Industrial sector detail: not included

Open or closed economy: NA

Capital markets: none

Retirement Plan Characteristics: no (planned)

Government behavior none

Federal budget: not represented

OASDI and HI trust funds: income and benefit payments, simple trust fund accounts

Regulations: none

Taxes: No tax system behaviors. Tax rules at individual person/family level.

Government employee pension programs none

Federal civil service: no

Military: no

State and local government, types: no

Public sector health care finance programs none

Medicare: no

Medicaid: no

Military/CHAMPUS: no

Veterans: no

Indian Health Service and others: no

Private sector health care finance programs none

Private health insurance, especially retiree health insurance: no (dental care insurance at person level only; no health plans)

Employer/plan sponsor behavior: no

Health care provider behavior: none

Insurer behavior: none

Institutionalization: no

Assumptions, Parameters, Methodology

Key Assumptions: Future behavioral characteristics will be like the past, in the sense that reduced form descriptive functions can continue to characterize relationships. Future estimates constrained to aggregate projections from the OASDI Trustees Report intermediate assumptions.

Types of Parameters, Decrements, Transition Rates/Probabilities

Experience considered, origins of decrements: Analysis and tabulation of public survey data bases. Econometric equations representing demographic and economic behavior estimated from large nationally representative longitudinal microdata bases (NLMS, PSID, NLS).

Consistency with other experience and other assumptions of model: All equations and operating characteristics estimated individually.

Internal consistency: Model is internally consistent. Relationships typically estimated as reduced form single equations.

Methodology used to estimate parameters and relationships

Econometric/statistical: OLS, logit, probit, tobit

Actuarial: Specification of program rules

Judgmental: yes

Economic/actuarial literature, studies done by others: yes

Simulation Methodology: Dynamic microsimulation, cross-section imputation

Stochastic Properties: Monte Carlo simulation. Rely on variance reduction techniques to minimize random variation.

Feedback Phenomena: No simultaneous feedbacks. Extensive feedback with one year lag.

Policy levers: Social security tax and benefit computation rules, tax rules. No pension rules or regulations.

Economic/demographic feedbacks

Employer costs and behavior: none

Labor market behavior: Labor force status affects some demographic behavior in following period. Assets, social security and employer pensions do not affect labor market behavior.

Taxes, government deficits, etc: none

Capital accumulation: Individual wealth accumulation is modeled, but no feedback on other behaviors.

Interest rates: exogenous (no feedback)

Employment, productivity, economic activity, GDP: exogenous

Sensitivity Analysis: Simulations can be done with alternative parameter values, but this is costly because of complexity of each simulation.

Model Validation Procedures: Considerable attention paid to historical alignment and validation. Simulations of 1960-1995 period compared in detail with historical data. Discrepancies used to estimate alignment parameters. Result is reasonably tight fit to detailed historical data, 1960-1995, and presumed basis for projecting alignment parameters.

Model Alignment Procedures: Discrepancies between summed micro outcomes and historical aggregates are used to estimate alignment parameters, which bring the summed micro outcomes into alignment with the historical aggregates. When the model is run into the future, the full set of time series of alignment parameters provides the basis for forecasting the alignment parameters into the future.

Model Debugging Procedures: An elaborate debugging apparatus automatically calculates simulated means for all input and output variables for each equation in the model, so that many bugs can be discovered and traced at the source. These extensive input-output calculations enable the user to observe the model carry out its detailed calculations, and are part of the routine output for each run.

Computer implementation

Hardware requirements	Dual and single PC's (primarily PentiumPro) up to 256 MB of RAM (16MB RAM required)
Software	C language; Linux operating system
Computer costs	Running costs marginal, virtually all runs done on Pentium microcomputers, as well as analyses of output
Transportability	Highly portable among computational platforms because of portability of C

Contact person: Steven Caldwell, Department of Sociology, Cornell University

ANNEX 5-2

CORSIM POLICY MATRICES

CORSIM -- POLICY MATRIX

1. Effects of Policy Measures on Employer Pensions

Outcome Variable	Offerings	Types of plans and provisions	Costs of plans	Funding	Contributions and benefits
Policy Input					
Tax Policy					
Pension					
General					
Social Security					
Retirement age					
Benefit structure					
Indexation					
Payroll tax					
Trust fund investment					
Individual accounts					
Funding and Guarantees					
PBGC premium					
Funding rules					
Pension Regulation and Policy					
ERISA/IRS					
Employer plans					
Pension and saving incentives/mandates					

Blank cell indicates that the effects of the policy issue or input on that outcome cannot be simulated in this model. CORSIM does not depict employer pensions.

CORSIM -- POLICY MATRIX

2. Effects of Policy Measures on Employees

Outcome Variable	Job availability	Portability	DC accumulations, investments, earnings	Benefit accruals ¹	Wage and non-wage compensation levels and mix	Incidence and timing of retirement
Policy Input						
Tax Policy						
Pension						
General						
Social Security						
Retirement age				x		
Benefit structure				x		
Indexation				x		
Payroll tax				x		
Trust fund investment						
Individual accounts						
Funding and Guarantees						
PBGC premium						
Funding rules						
Pension Regulation and Policy						
ERISA/IRS						
Employer plans						
Pension and saving incentives/mandates						

Blank cell indicates that the effects of the policy issue or input on that outcome cannot be simulated in this model.

¹ Social Security benefits only. CORSIM does not depict employer pensions.

CORSIM -- POLICY MATRIX

3. Effects of Policy Measures on Retirees

Outcome Variable	Payouts	Funded levels of plans	Retirement income	Replacement rates	Poverty levels	Health care costs and insurance	Retirement age and labor mkt outcomes	Inflation protection	Auxiliary benefits
Policy Input									
Tax Policy									
Pension									
General									
Social Security									
Retirement age			x	x	x				
Benefit structure			x	x	x				
Indexation			x	x	x				
Payroll tax			x	x	x				
Trust fund investment									
Individual accounts			x	x	x				
Funding and Guarantees									
PBGC premium									
Funding rules									
Pension Regulation and Policy									
ERISA/IRS									
Employer plans									
Pension and saving incentives/mandates									

Blank cell indicates that the effects of the policy issue or input on that outcome cannot be simulated in this model.
 CORSIM does not depict employer pensions.

CORSIM -- POLICY MATRIX

4. Effects of Policy Measures on Industry Outcomes

Outcome Variable	Financial strength of plans, sponsors, insurers	Labor costs	Profits	Competitiveness
Policy Input				
Tax Policy				
Pension				
General				
Social Security				
Retirement age				
Benefit structure				
Indexation				
Payroll tax				
Trust fund investment				
Individual accounts				
Funding and Guarantees				
PBGC premium				
Funding rules				
Pension Regulation and Policy				
ERISA/IRS				
Employer plans				
Pension and saving incentives/mandates				

Blank cell indicates that the effects of the policy issue or input on that outcome cannot be simulated in this model.

Industries are not modeled in CORSIM. CORSIM shows effects only on individuals.

CORSIM -- POLICY MATRIX

5. Effects of Policy Measures on Aggregate Economy

Outcome Variable	GDP growth	Saving and capital accumulation	Equity investment	Investment efficiency	Interest rates	Productivity	Inflation	Labor mobility and labor market flexibility
Policy Input								
Tax Policy								
Pension								
General								
Social Security								
Retirement age								
Benefit structure								
Indexation								
Payroll tax								
Trust fund investment								
Individual accounts								
Funding and Guarantees								
PBGC premium								
Funding rules								
Pension Regulation and Policy								
ERISA/IRS								
Employer plans								
Pension and saving incentives/mandates								

Blank cell indicates that the effects of the policy issue or input on that outcome cannot be simulated in this model.

Aggregate economy is not modeled in CORSIM.

CORSIM -- POLICY MATRIX

6. Effects of Policy Measures on Government Finances

Outcome Variable	Tax revenue	Expenditures by program	Deficits and debt	Social Security and Medicare
Policy Input				
Tax Policy				
Pension				
General	x			
Social Security				
Retirement age		x		x ¹
Benefit structure		x		x ¹
Indexation		x		x ¹
Payroll tax	x			
Trust fund investment				
Individual accounts		x		
Funding and Guarantees				
PBGC premium				
Funding rules				
Pension Regulation and Policy				
ERISA/IRS				
Employer plans				
Pension and saving incentives/mandates				

Blank cell indicates that the effects of the policy issue or input on that outcome cannot be simulated in this model.

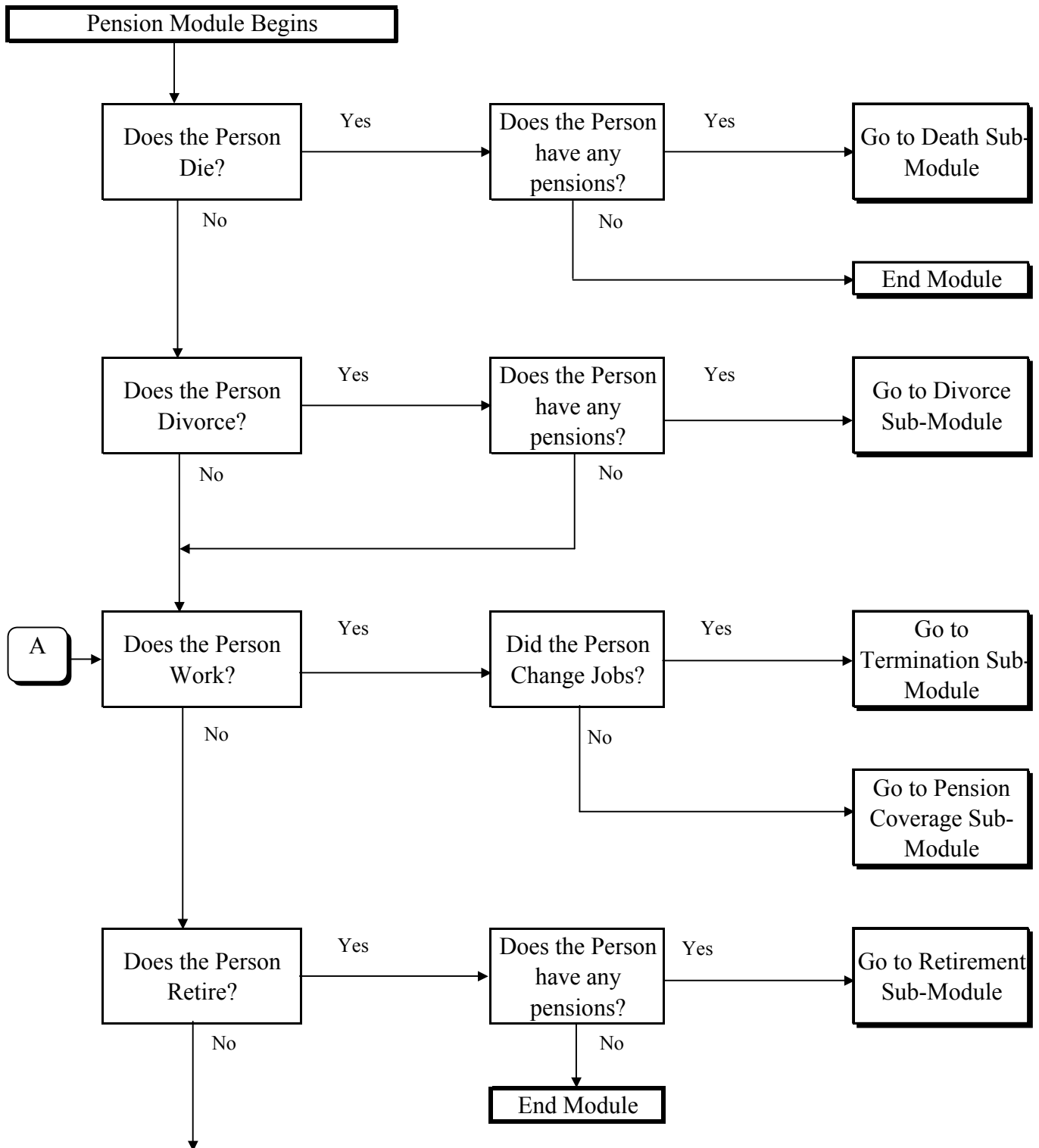
¹ Social security benefit payments only.

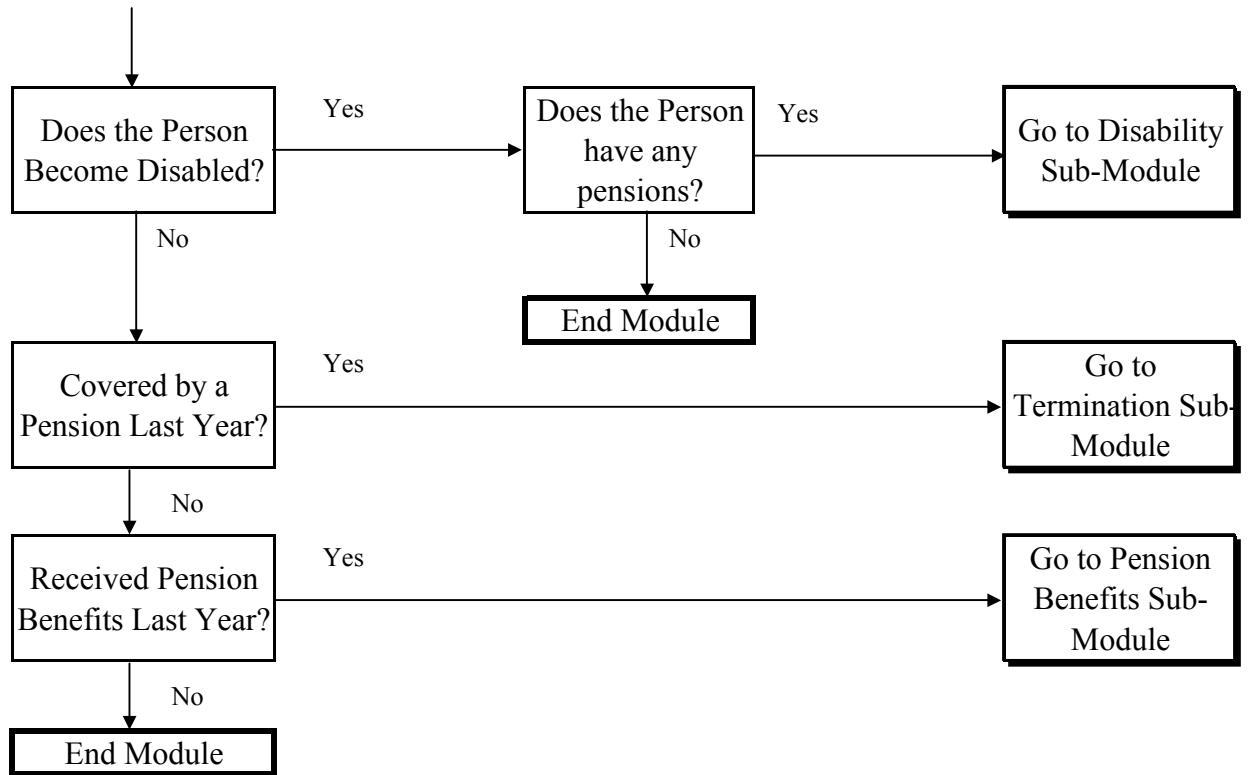
ANNEX 5-3

PROPOSED CORSIM PRIVATE PENSION MODULE FLOWCHARTS

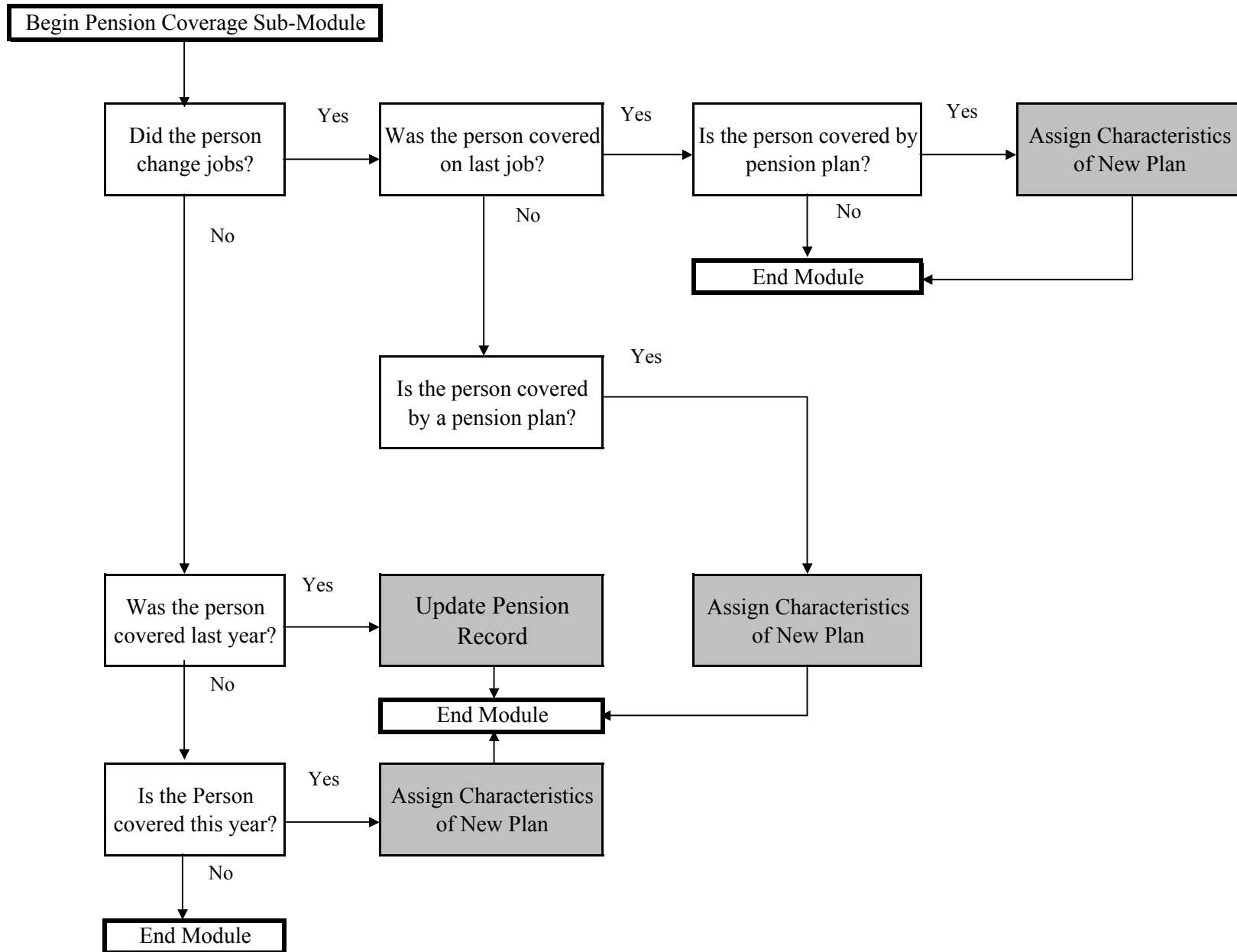
The following flowcharts show the proposed structure of the private pension module to be added to CORSIM in 1998. The flowcharts were developed by Sonya Williams of Cornell University.

Private Pension Main Module

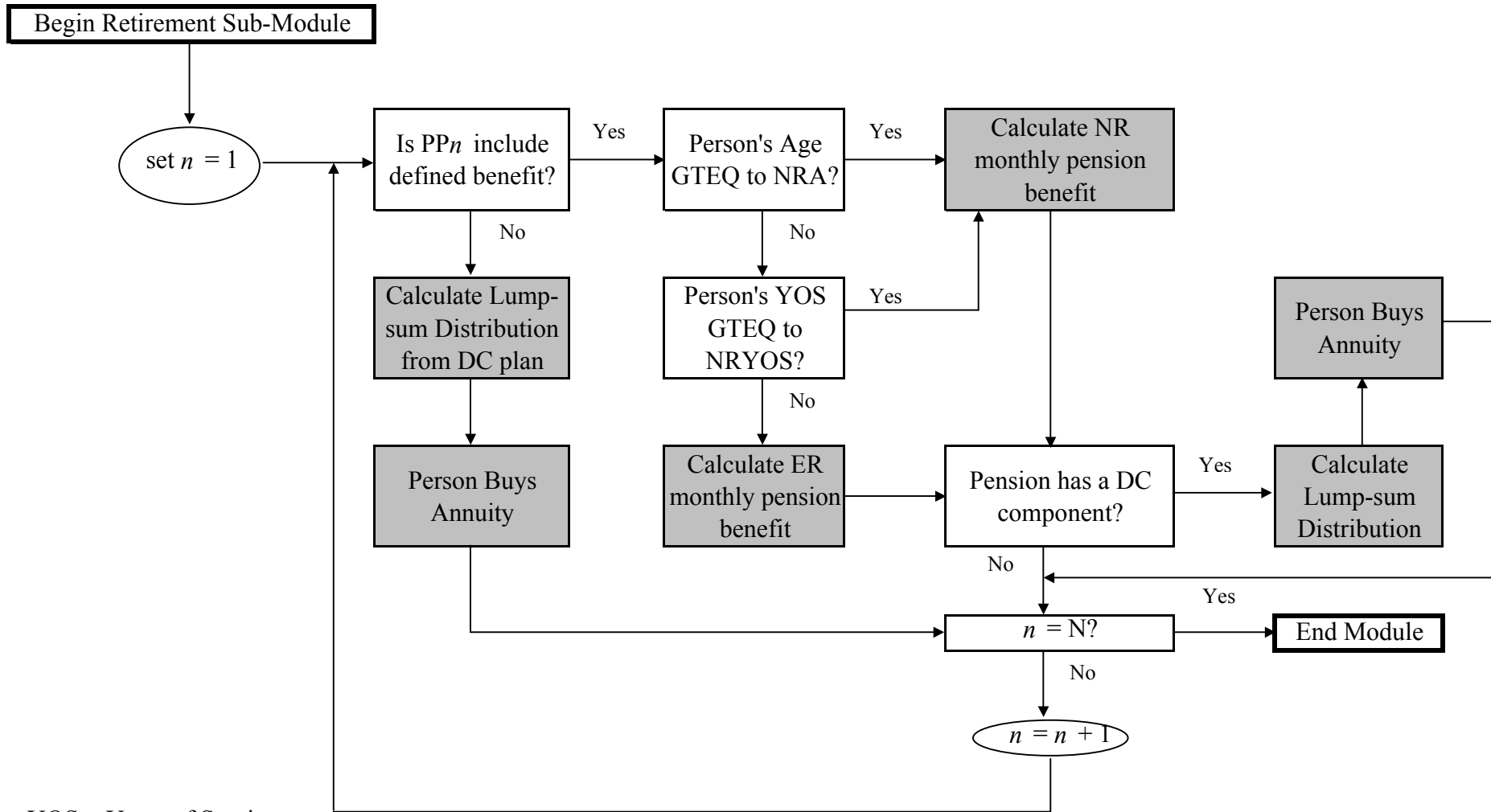




Pension Coverage Sub-Module

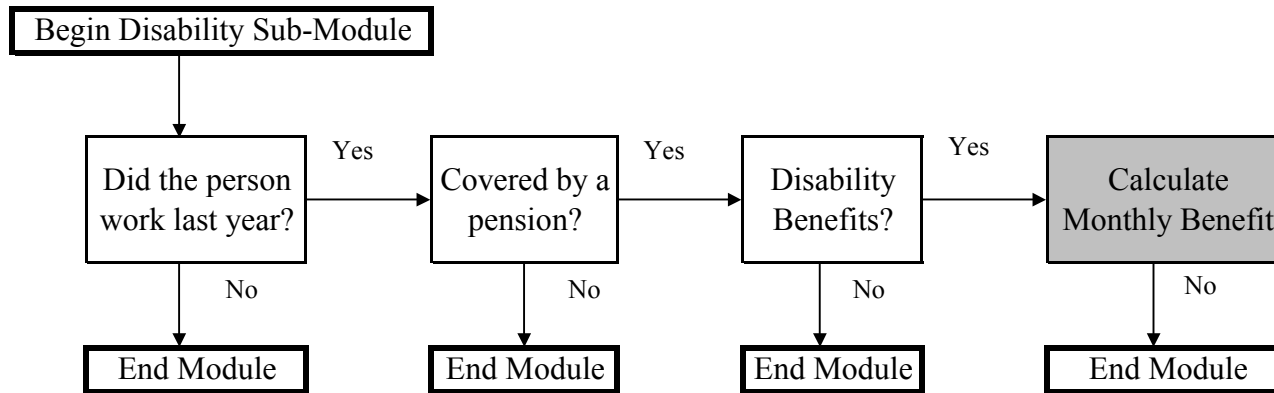


Retirement Sub-Module

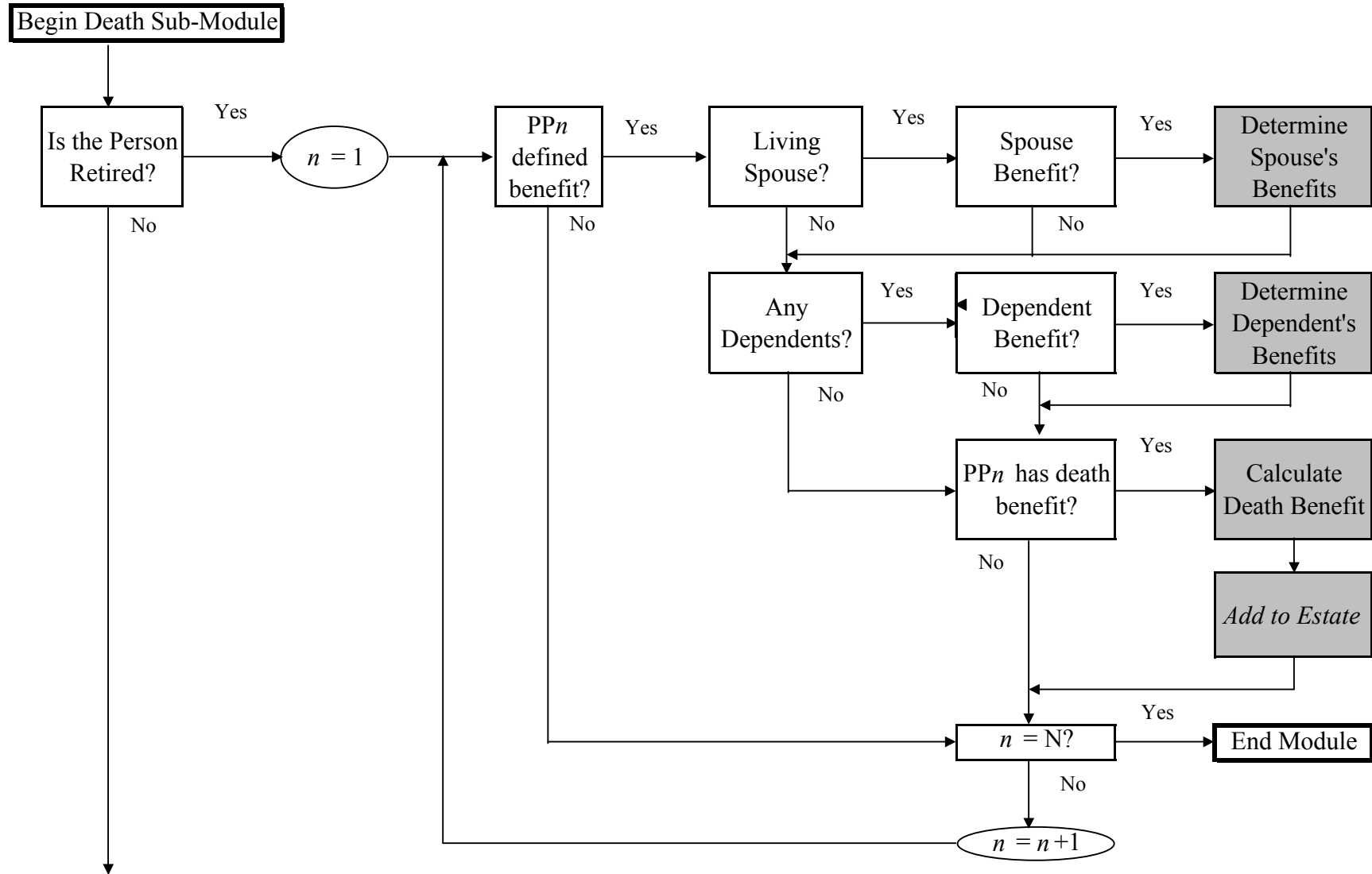


YOS = Years of Service
 GTEQ = Greater than or Equal to
 NR = Normal Retirement
 ER = Early Retirement

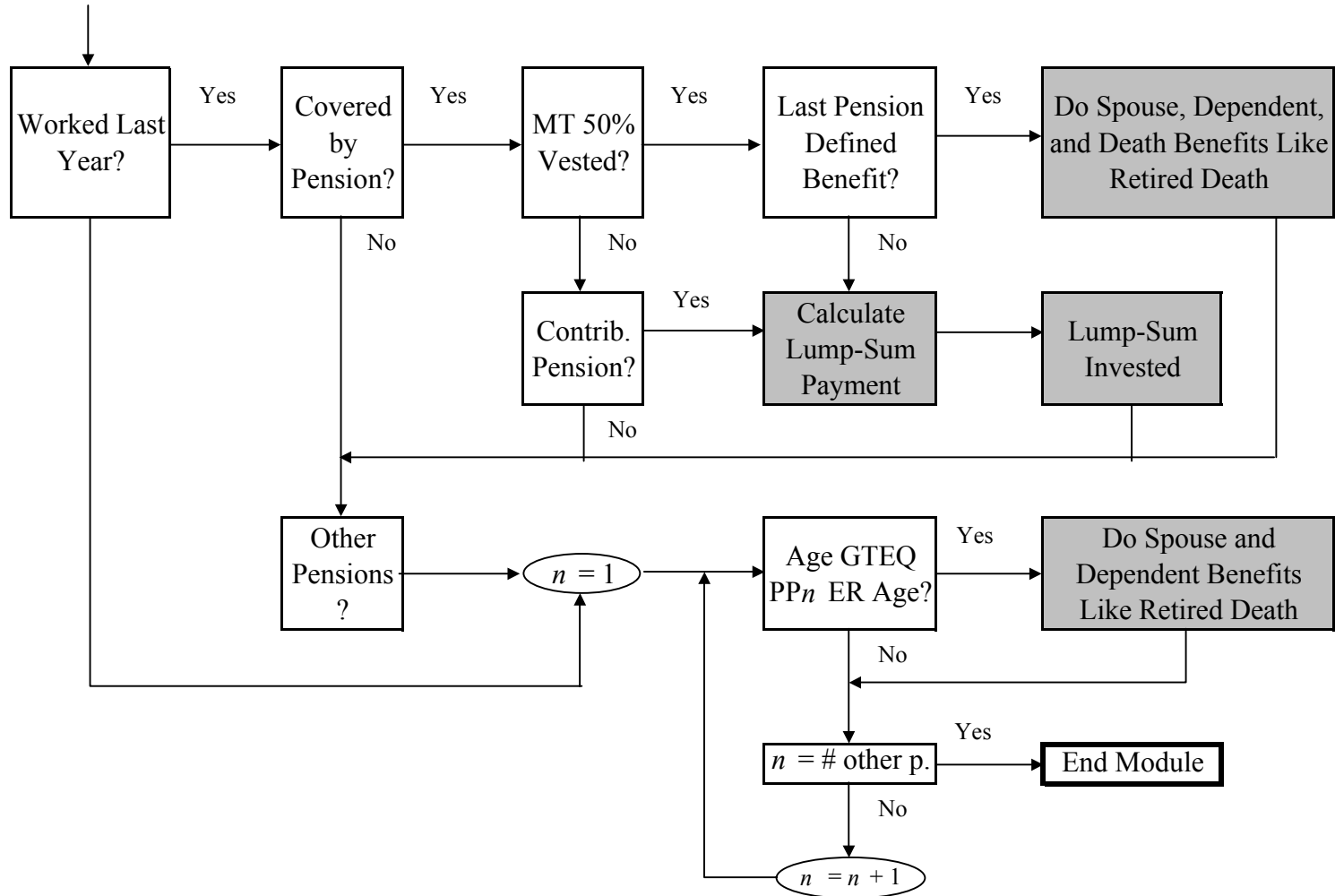
Disability Sub-Module



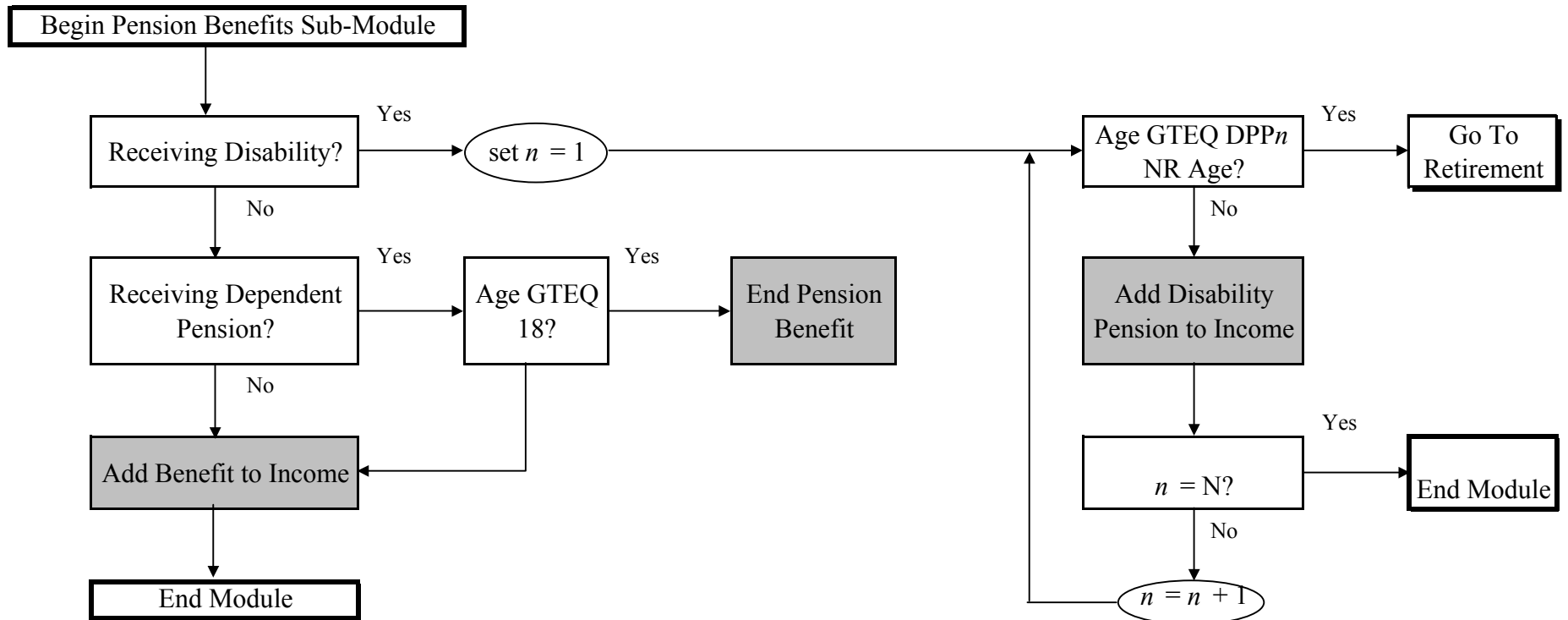
Death Sub-Module



Death Sub-Module

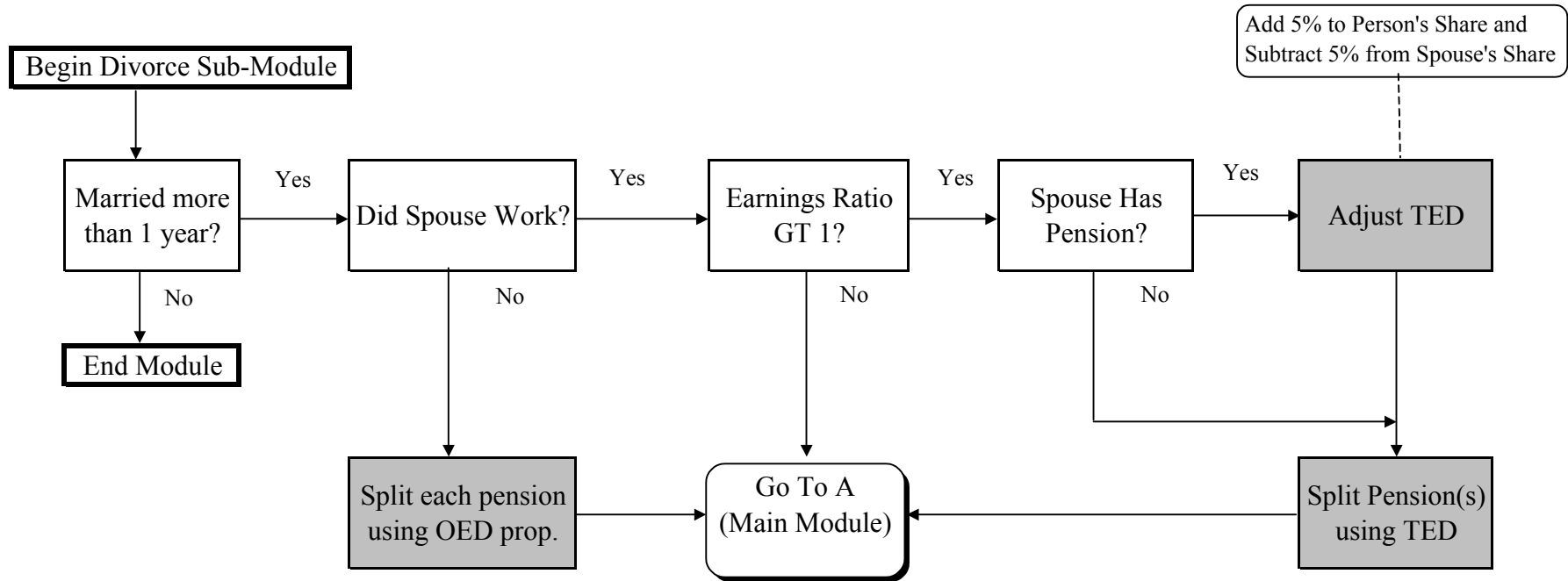


Pension Benefits Sub-Module



DP_n = Pension Plan from which person receives disability benefits

Divorce Sub-Module

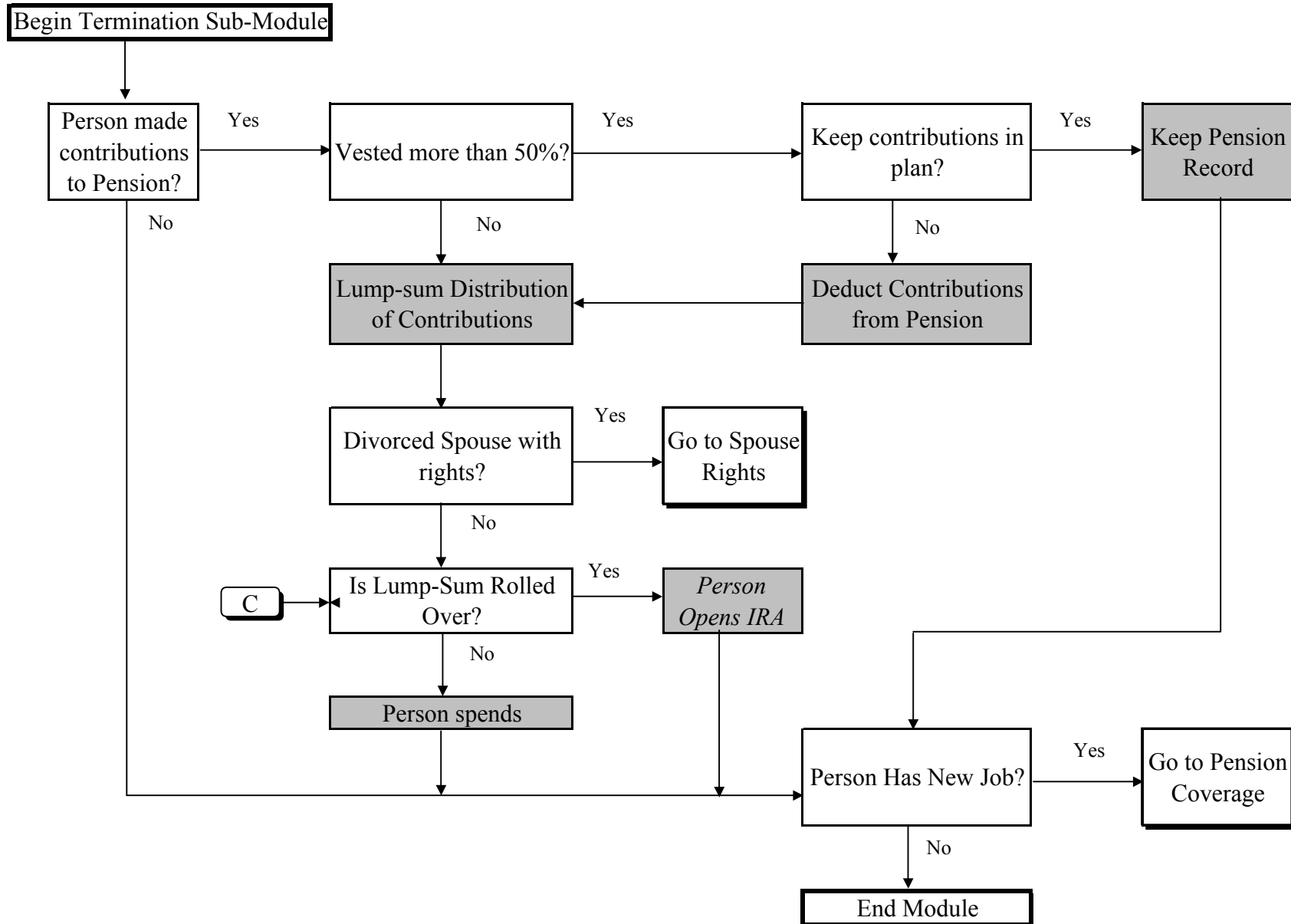


Earnings Ratio = Person's Earnings / Spouse's Earnings

One Earner Divorce (OED) Proportions		
<u>Years Married</u>	<u>Person Share</u>	<u>Spouse Share</u>
2 - 3	75%	25%
4 - 5	70%	30%
6 - 7	65%	35%
8 - 9	60%	40%
>= 10	50%	50%

Two Earner Divorce (TED) Proportions		
<u>Earnings Ratio</u>	<u>Person Share</u>	<u>Spouse Share</u>
>1, <2	65%	35%
>=2, <3	60%	40%
>=3, <=4	55%	45%
>= 5	50%	50%

Termination Sub-Module



Termination Sub-Module

