Overview of

DYNACAN

a full-fledged Canadian actuarial stochastic model designed for the fiscal and policy analysis of social security schemes

written by Rick Morrison (November 1998)
slightly adapted by Bernard Dussault (March 2000) for inclusion on the IAA website

1.1   Executive Summary

DYNACAN is a **Canadian**, **stochastic** (Monte Carlo), **open**, **longitudinal**, **dynamic microsimulation** model developed for generating longitudinal and **cross-sectional**, as well as **fiscal** and **policy-oriented analyses** of Canadian Social Security Schemes such as the Canada Pension Plan (CPP), the Old Age Security (OAS) Program and the Guaranteed Income Supplement (GIS). DYNACAN is validated in many respects, the main one being that its demographic and financial results are made consistent with those of the macrosimulation model ACTUCAN. ACTUCAN is the model used by the **Chief Actuary for Canadian public pensions plans** for purposes of the statutory actuarial valuations of the CPP and the OAS.
-        DYNACAN is a <u>microsimulation</u> model in the sense that its fundamental unit of analysis is the individual. However, these individuals, over 212,000 of them in the full version of the starting database, are always treated in the context of the families in which they reside.
-        It is <u>Canadian</u> in that, collectively, the individuals are generally representative of the total Canadian population. This representativeness applies first in 1971, for the starting data base, and then on through the projection period as the modelled population evolves as a result of the simulated events. However, the population of Quebec, though modelled in DYNACAN, is typically not directly relevant for CPP calculations.
-        DYNACAN is <u>longitudinal</u> in the sense of projecting the individuals forward through time, synthesizing entire realistic life histories as regards the events relevant for CPP contributions and benefits. This is appropriate given its companion status to the CPP Actuarial Valuation Model (ACTUCAN).
-        It is <u>dynamic</u> in the sense of generating an evolving population, whose characteristics -- size, membership and family configurations -- change over time.
-        It is a <u>stochastic</u> (Monte Carlo) model in the sense of using "random" numbers in its simulation of the events (births, deaths, marriages, retirement, etc.) associated with that evolution.
-        DYNACAN is <u>open</u> in the sense that it allows additions to the population through births and immigration and departures through deaths and emigration.

The DYNACAN model is a collection of programs that together permit the longitudinal simulation of Canadian **demographics**, **earnings** and **other characteristics** sufficient for the calculation of annual Canada Pension Plan (CPP) contributions and benefits at the level of the individual and the family. The model's projections/simulations typically extend over multiple decades ranging from the past into the future, e.g. from 1971 to 2030. The primary purpose of DYNACAN is to enable one to project both the evolution of the CPP as it currently stands, and the consequences of hypothetical or proposed changes to the CPP. The latter usage includes the projection of detailed distributional impacts, and, where appropriate, winners/losers analyses, on into the future.

The development of DYNACAN through the autumn of 1998 has taken place under the direction of the Office of the Chief Actuary in Canada's Office of the Superintendent of Financial Institutions (OSFI). From that point on, model governance will progressively transfer to Human Resources Development Canada (HRDC), primary user of the model's results. Funding for DYNACAN has proceeded on a cost-recovery basis from the Canada Pension Plan, with budgets approved annually in advance by a Project Board consisting of the Chief Actuary together with senior HRDC officials including the CPP Comptroller. HRDC is responsible for policy development and for the financial and operational management of the CPP. The choice of OSFI as DYNACAN's organizational home reflected the desire to maintain **consistency between results from DYNACAN and those from the CPP Actuarial Valuation Model (ACTUCAN)** that is maintained by the Chief Actuary in support of statutory reporting requirements for the CPP. Now that the model has become operational, control will increasingly reflect its role as a policy model used by HRDC. However, if the Canadian government is to speak with a single voice on CPP policy, there will need to be an intimate, continuing co-ordination with the ACTUCAN model as the result of DYNACAN's dependence on ACTUCAN for several critical aspects of CPP projection. Included among these are the calculations of CPP contribution rates, the projection of the CPP Account and Fund, and the desire for common assumptions. As well, clients have indicated their desire to assess the degree of macro-level agreement between DYNACAN and ACTUCAN on policy option impact projections on an option by option basis.

Historically, in the first half of the 1990s, following an identification of anticipated CPP policy modelling needs, Canada conducted a study of international microsimulation practices. Seeking to avoid reinventing the wheel, the project team elected to adopt a functional, existing longitudinal microsimulation model as a template for developing DYNACAN. Investigations showed that **the CORSIM model, developed at Cornell University (New-York State, US) under the leadership of Professor Steven Caldwell**, best matched the team's requirements. In practice, not only has the original CORSIM model source code proved extremely valuable, but DYNACAN has also benefitted considerably from a subsequent ongoing collaboration with Professor Caldwell and his CORSIM team.

The primary DYNACAN computing environment consists, in fall 1998, of a dual-processor 200 MHz Pentium Pro machine running parallelized code, 512 megabytes of RAM, a fast hard disk and a writable CD-ROM capacity to meet the very heavy computational and I/O demands associated with dynamic microsimulation. Most of the model's components are built in the C language, and run under the Linux operating system. Other components, e.g. programs like SAS and TPL used for post-processing, secondary analysis, and parameter estimation, use the Windows NT 4.0 operating system, or a DOS emulator running under Linux.

A critical project requirement is that DYNACAN's projections be consistent with those of OSFI's CPP Actuarial Valuation Model, ACTUCAN, when comparable assumptions are used. In practice, this requirement means inter alia that the rates, e.g. mortality rates, associated with DYNACAN's projected flows are aligned to those of ACTUCAN. DYNACAN is used to ascertain the financial impacts on individuals/families of CPP policy proposals, once such proposals have been analyzed, primarily in aggregate, by OSFI's Valuation Unit.

This document seeks generally to parallel Appendix B of the statutory actuarial reports produced by the Chief Actuary. It emphasizes DYNACAN's methodology, while also describing the underlying assumptions and the data sources employed. It provides an overview of the operational substance of DYNACAN, serving, as its title indicates, as high-level technical documentation for DYNACAN. Although the documentation is somewhat technical in nature, it is written to be high-level in the sense of describing the methodology for the whole of the DYNACAN package. By explicit intention, it does not provide detailed technical descriptions of individual modules or aspects, or the statistical or programming details. It makes frequent use of references to lower-level technical documentation that has been prepared for various aspects of DYNACAN.

1.2 High level documentation on DYNACAN

(a) Overview: At its most basic level, a DYNACAN analysis, or "computer run," simulates, for a representative sample of individual Canadians, over a given sequence of years, all of the CPP-relevant events in which they are, or were, involved, all of the contributions they pay, and the benefits they receive. "CPP-relevant" is used in a broad sense that includes, for example, education, since educational attainment affects both the average level and variability of earnings, which in turn affect CPP benefits. For typical impact analyses, the simulation examines the differences between two parallel systems, a "base" system that is usually the current system, and an option system that represents some change being considered for the CPP. This high level documentation provides a very high-level characterization of the substantive mechanics of such an analysis; subsequent sections in this high-level technical documentation then begin to fill in the myriad of details. As noted, however, the lower-level details are left to programmer-level technical documentation, and to the many in-line comments contained in DYNACAN's source code.

(b) General Methodology for Demographic and Economic Projection: For a typical multi-year DYNACAN analysis, during each simulation year DYNACAN initiates, via Monte Carlo simulation, appropriate events for each person in the sample, beginning with some 200,000 individuals for the full sample in 1971. Year by year, realistic scenarios for the representative individuals are incrementally extended in a fashion consistent with historical experience and the user's assumptions. Thus, simulation year by simulation year, the total and the mix of constituent individuals in the sample change because new

individuals are born or immigrate, while others die or emigrate. Of course, those who remain may reconfigure themselves as regards the families they live in. Generally speaking, the sample size itself grows substantially throughout the course of the simulation as a result of these events, as newborns and immigrants are added to the data base, while the information on deceased individuals and emigrants is retained.

☞   Demographically, certain women in the model become pregnant and give birth, with probabilities consistent with their ages, marital, education and employment statuses, and the number of previous children. Similarly, individuals die, with probabilities consistent with their configuration of gender, age, education, and with their marital, employment and disability status, as well as with mortality improvements over time. Some individuals improve their educational attainments. Some individuals leave home and/or get married, thereby creating new households. Still other individuals become disabled, or recover from disabilities. New families immigrate to Canada, and some Canadian families emigrate. Within Canada, families may move between Quebec and the rest of Canada.

☞   A similar situation applies to the economic side of the individuals' lives. Each simulation year, some individuals enter or leave the labour force. Some workers remain employed, while others experience varying degrees of unemployment. Those who are working receive employment earnings that are not only generally consistent with characteristics such as age, gender and education and full-time/part-time status and full-year/part-year status, but that also reflect appropriate variations across persons and over time. Their employment earnings may lead them to contribute to the CPP or the QPP, and these contributions will establish entitlements to CPP benefits when the other triggering conditions are met.

DYNACAN simulates all of these demographic and economic events via Monte Carlo simulation using probabilities generated to be individually consistent with the sample individuals' personal and/or family characteristics. At the same time, the probabilities are collectively constrained to be consistent with administrative, census and survey data, the results of published research, user assumptions, and the assumptions of the CPP Actuarial Valuation Model, ACTUCAN.

☞   Finally, DYNACAN calculates, for each individual in the sample, and for each year in the simulation, the CPP contribution/benefit implications of the various life events. One of DYNACAN's defining characteristics is that, for the base system these benefits are calculated, both initially and in terms of their subsequent year to year indexation, via very close approximations to the same formulae that the CPP uses to administer its benefits. When an application of DYNACAN develops comparative impact analyses, the second, option, system will reflect whatever alternative CPP calculations the clients deem appropriate. The reader should be aware that DYNACAN's design also permits, in principle, feedback in the sense of allowing for the differing incentives of the option system to result in

the occurrence of different life and labour force events and different CPP contributions and benefits. However, the practical exigencies of allowing this kind of feedback ensure that it is seldom implemented.
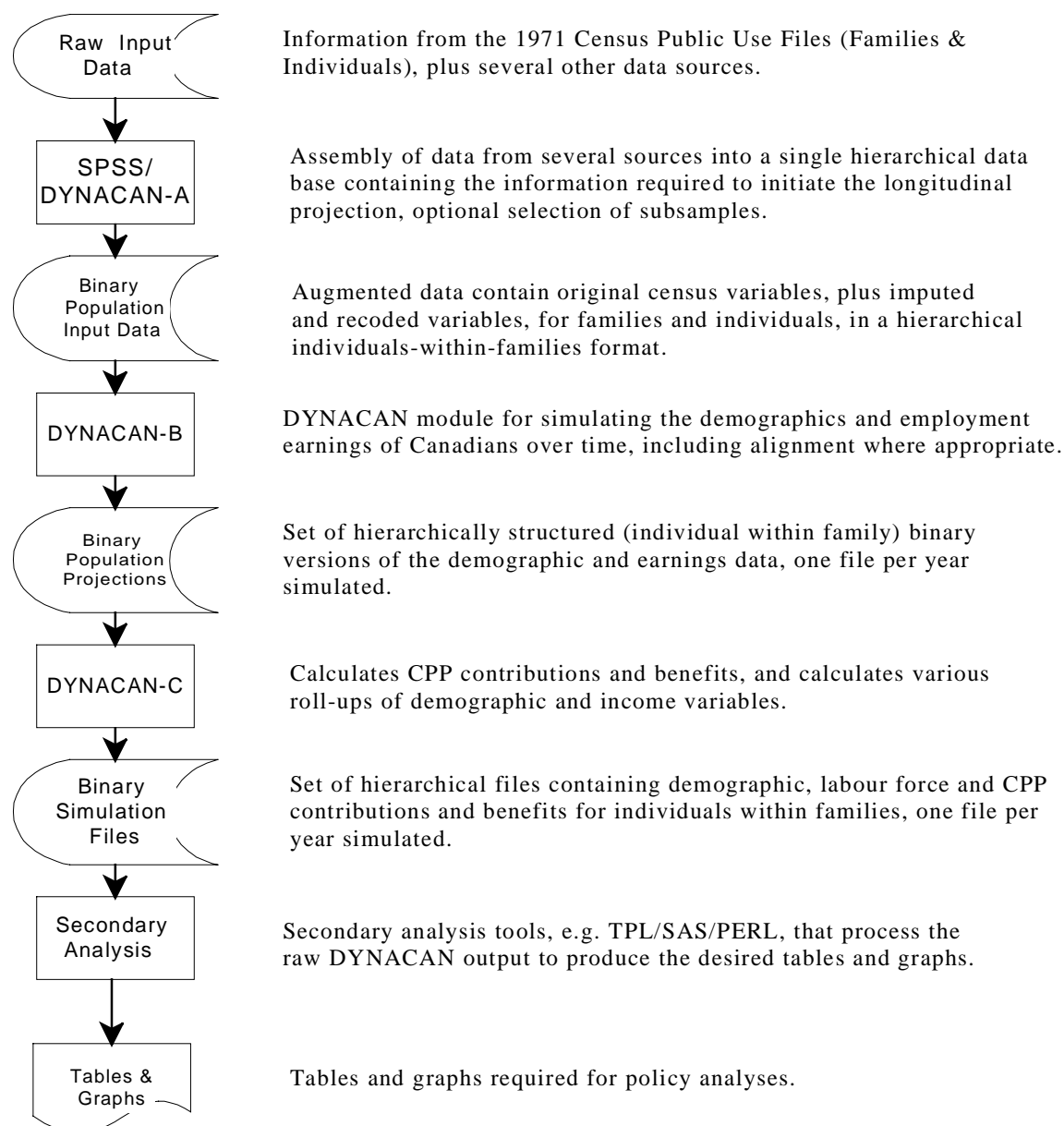
(c) Usage of DYNACAN's outputs: A major strength of the microsimulation approach is that an analyst can then assemble relevant portions of the model's micro-level information in whatever manner best addresses the policy issue at hand. The analyst can select data from some individuals, but not others, or from certain periods, but not others. The analyst can compare information between one simulation of the CPP, the current system, and an alternative simulation in which the calculations are altered to reflect some proposed policy, to assess policy impacts and to identify prospective winners and losers. Finally, of course, the analyst can collate and present the relevant data and the impacts in whatever fashion best suits the goals of the analysis, or the target audience.

Although the preceding description is faithful to the substance of the events and calculations included in DYNACAN, it does not effectively capture the sequence of calculations performed during an analysis. Figure 1-1 (page 6), A DYNACAN Overview," better represents the model's internal structure, though still only at a very high level. From the perspective of its operational substance, DYNACAN has three main components –
- DYNACAN-A (**A**ssemble initial representative data base),
- DYNACAN-B (**B**asic demography and economics) and
- DYNACAN-C (**C**alculate CPP Contributions and Benefits).

Methodologically, these three components perform the calculations described below:

Figure 1-1: DYNACAN Overview

Flow	Description
Raw Input Data	Information from the 1971 Census Public Use Files (Families & Individuals), plus several other data sources.
SPSS/ DYNACAN-A	Assembly of data from several sources into a single hierarchical data base containing the information required to initiate the longitudinal projection, optional selection of subsamples.
Binary Population Input Data	Augmented data contain original census variables, plus imputed and recoded variables, for families and individuals, in a hierarchical individuals-within-families format.
DYNACAN-B	DYNACAN module for simulating the demographics and employment earnings of Canadians over time, including alignment where appropriate.
Binary Population Projections	Set of hierarchically structured (individual within family) binary versions of the demographic and earnings data, one file per year simulated.
DYNACAN-C	Calculates CPP contributions and benefits, and calculates various roll-ups of demographic and income variables.
Binary Simulation Files	Set of hierarchical files containing demographic, labour force and CPP contributions and benefits for individuals within families, one file per year simulated.
Secondary Analysis	Secondary analysis tools, e.g. TPL/SAS/PERL, that process the raw DYNACAN output to produce the desired tables and graphs.
Tables & Graphs	Tables and graphs required for policy analyses.

RJM, DC-OVR.CFL, 1998/11/03

DYNACAN-A (**A**ssemble Initial Representative Data Base): The starting point for
DYNACAN's input data base, in terms of both demographics and the initial distribution
of employment earnings, is Statistics Canada's Public Use Microdata Files for the 1971
census. First, the demographic and employment earnings information for individuals from

the Individuals file is merged with that for individuals appearing in the Families file (reflecting those individuals living in families of two or more persons). The resulting file consists of individuals grouped within family units. Unattached individuals are treated, from here on, as constituting legitimate "one-person families" in which the single member is the family head. Such treatment is identical to that used in cross-sectional microsimulation packages such as SIMTAB or Statistics Canada's SPSD/M. The data base is treated as generally representative of the Canadian population and its distribution as of June 30, 1971, with earnings corresponding to the full 1970 calendar year. However, to meet the requirements for use as the starting point in longitudinal microsimulation, the data require several adjustments. We summarize them here.

Some variables present in the Families file are absent from the Individuals file because they were unnecessary there, e.g. a code for the head of the household, and fields for the numbers of persons in the census or economic or census family. These variables were added to the records coming from the Individuals file. Similarly, several codes were inconsistent between the data from the Individuals and Families files; these were reconciled to produce the required consistency. Further, other variables in the census data were recoded to make them more useful, e.g. adding 1 to the variable for the number of persons present in the family, so that it reflects the total number of persons, rather than just the number of persons other than the head. Such changes were not required in a strictly logical sense, but they reduce the potential for errors when one later uses the data base. Other changes were slightly more ambitious, e.g. imputing specific ages and genders to individual children who, in the Families file, were simply noted as falling within a particular broader age grouping. Specific ages, genders and birth dates were synthesized for children aged 25 or older and living at home.

After these adjustments were complete, the resulting database was reasonably representative of the Canadian population, but still required some changes to be suitable as an input to DYNACAN-B. One such adjustment involved smoothing the age distributions for men and women. The extraction of the 1% public-use sample had yielded, as the inevitable consequence of sample variation, gender-specific, age-distributional curves that were more jagged than the known overall distributions by age. Further, a truncation of ages at age 75 for many of the individuals appearing in the sample yielded an inappropriate spike in the population at age 75, and a relative scarcity of individuals above that age. Important, but fairly minor, adjustments to the ages of selected individuals in the sample, and in a very few cases their genders, plus the cloning of a few individuals aged 75 and older, produced appropriately smoothed distributions that reflected known aggregates. Similar, but even less intrusive, change-of-region adjustments ensured that the age distributions for the sub-populations in Quebec and the rest-of-Canada (RoC) were also sufficiently smooth. Taken together, these adjustments of gender, age and region affected less than one percent of the individuals in the merged sample.

Other adjustments were necessary in DYNACAN-A to provide the variables that DYNACAN-B requires generating appropriate life events for the individuals. By way of non-exhaustive examples,

☞   Employment earnings that were known, collectively, to be earned by persons other than the head and spouse in a census family, had to be allocated to specific family members (children).

☞   Because the CPP began operation in 1966, and CPP benefits are dependent on the contributors' full contribution histories, the sample individuals' earnings histories had to be "backfilled" for the years 1966 through 1969. These histories are also potentially relevant for subsequently establishing eligibility for CPP Disability benefits.

☞   Disability status had to be imputed as of June 1971, since disabled individuals would not generally be employed. Neither would they be making CPP contributions that would establish entitlements to subsequent CPP benefits.

☞   To ensure conformity with survey results involving the proportions of women who report themselves as having never worked, certain women, with appropriate characteristics, had to be labeled as "excluded from entering the labour force." Otherwise, relative to survey and administrative data, too many of these women would subsequently be simulated as establishing CPP entitlements in their own right.

☞   The educational attainment variables available in the census had to be converted to the specific variables and codes used by DYNACAN-B's education module, and used in its earnings equations.

At the conclusion of these adjustments, the augmented census data exists in a binary population file suitable for input into DYNACAN-B. The full version of that file contains records for some 212,935 individuals, grouped into census families. However, a sample sub-file, consisting of 20% of the full sample's families and their constituent individuals, has been independently smoothed as to its Quebec and rest-of-Canada, gender-specific age distributions. A related sample, consisting of 40% of the full sample has also been created. The sample sub-files see regular use in analyses, validations and tests for which the sample size is not critical.

DYNACAN-B (**B**asic Demography and Employment Earnings): DYNACAN-B simulates the demographic and economic events relevant for the CPP. Given the complexity of this program, the description here divides the material into a set of relatively self-contained "nested" topics –

- A) Presentation of DYNACAN-B as a process that ages the underlying data base by one year, and that is run repeatedly to generate the multi-year demographic evolution needed for the whole simulation,
- B) Identification of the set of individual modules that make up the DYNACAN-B process, each of them simulating an event relevant to the demographic evolution,
- C) The internal structure of a typical, probability-oriented, DYNACAN-B module, including the role of alignment so that the expected numbers of events are constrained to respect the client's assumptions about event probabilities.
- D) High-level outlines of the several DYNACAN-B modules.

Experience in describing DYNACAN to a variety of audiences suggests that this top-down approach, despite a certain degree of redundancy, is the most effective way of communicating the DYNACAN-B structure.

A) DYNACAN-B as a repeated process: DYNACAN-B ages a representative population across time. Its "kernel" ages an input database by one year in any given pass. During each such pass, it simulates all of the births, deaths, marriages, labour force entry and exit and earnings, etc., that occur during that simulation year, and ages each of the individuals in the database by one year. It then outputs another database that is itself a new, representative population, but one that reflects the situation one year later than did the previous input database. The cycle is repeated over and over again for the length of the simulation run; in each cycle, the output data base from one pass through the kernel is used as the input for the next pass. Throughout these calculations, DYNACAN usually retains all of the evolving population in memory.

A typical run will consist of 60 or more passes, corresponding to an equal number of simulation years. The collective output from the DYNACAN-B stage is thus a sequence of cross-sectional databases, one per simulated year. The key characteristic for the outputs' collective suitability for subsequent longitudinal analyses is that each individual has a unique personal identifier; that identifier is included in each of the sequence of year-specific data bases generated as the simulated individuals age between one data base to the next. (We emphasize that the DYNACAN-generated personal identifiers used in the model for these synthetic individuals are not in any way related to the Social Insurance Numbers for the individuals in the starting population. The Census Public Use Microdata files that provide DYNACAN's starting base do not include Social Insurance Numbers.)

Using the collation tools available in DYNACAN, one can conveniently track specific individuals and the events that apply to them over time. Although family structures are more volatile than individuals, DYNACAN also provides each family unit with a unique identifier. The analysis tools, including the use of several standard Unix utilities operating under Linux, allow one to track a family from one year to the next. The earlier one-paragraph characterization of DYNACAN-B has already conveyed the notion of a "kernel" that repeatedly ages a representative data base one year at a time,

through all of the years of the simulation. Here, it is appropriate to give a better idea of what the internal structure of that kernel looks like.

B) The Aging Process as a Sequence of Modules: DYNACAN-B consists of a number of modules executed in sequence, each of them modifying the in-memory population for that module's event for the current year. Each module processes all of the population for which that module/event is relevant, updating that aspect of the individuals' lives. Clearly, not all individuals are eligible for all modules; e.g. individuals who have previously died will not give birth, and individuals who are presently married are not, in the same year, eligible to enter the marriage market.

Once the full set of modules has been executed, they have collectively aged the in-memory population by one year. That is, they have implemented all of the events that effectively transform the base from one year's representative population to the next year's representative population. Table 1-1 below indicates the order of DYNACAN-B's modules.

TABLE 1-1: ORDER OF MODULE EXECUTION IN DYNACAN-B

1) Emigration (from Canada)
2) Pseudo-immigrants/emigrants (both into Canada and leaving Canada – serving as a reconciliation between Statistics Canada's treatments of population stocks and population flows; includes previous emigrants returning to Canada)
3) Internal Migration (bi-directional between Quebec and the Rest of Canada)
4) Immigration (to Canada)
5) Mortality
6) Fertility
7) Aging (incrementing by one year the ages of those who have survived)
8) Marriage (including remarriage) – with two sub-modules:
   a) Entry of men and women into the pool of those seeking to become married
   b) Marriage market (matching specific men and women as couples)
9) Divorce (including termination of common-law unions)
10) Leaving Home (departure of children relevant for family formation)
11) Education (including later-life education)
12) Disability for CPP purposes (including rehabilitation)
13) Marking certain persons as never working
14) Retirement (for CPP retirement benefit purposes only, includes early retirement; post CPP-retirement employment is still possible)
15) Earnings -- with three major sub-modules:
   a) Labour force participation, including paid versus self-employment status
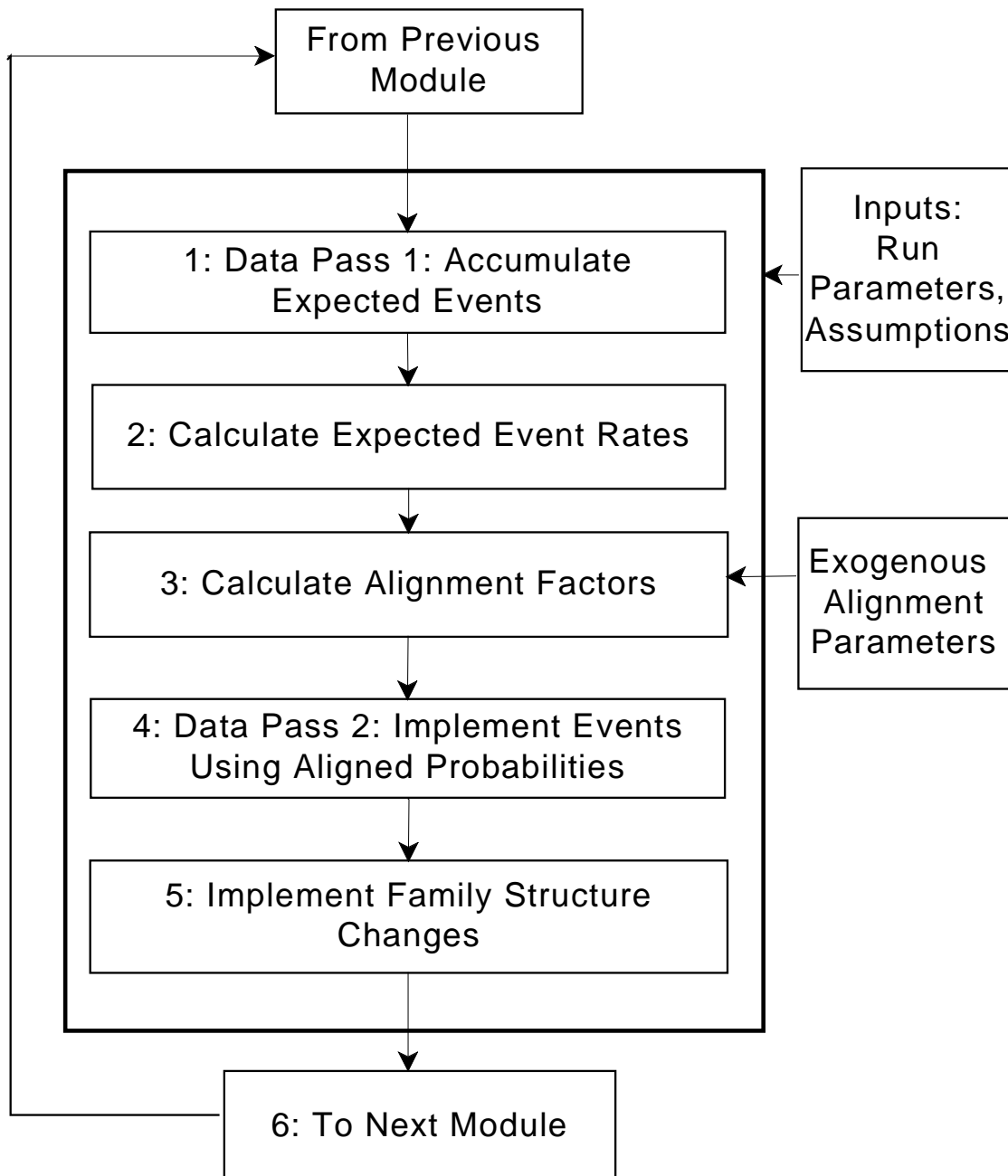   b) Total Wages

c) Weeks worked

In general, the implementation of DYNACAN-B for a single simulation year in a simulation run means executing all of the modules, in the order shown in Table 1-1. Each such execution often requires two passes through the database as it then exists in the computer's memory. The next section provides a slightly fuller characterization of the several steps for a typical probability-oriented module, i.e. one that determines whether a particular event occurs for a specific individual.

C) The Structure of a Typical Module: For purposes of this high-level characterization, the internal structure for a typical module can be thought of in terms of five major steps. We'll indicate the steps here via a summary graphic, and then provide slightly fuller descriptions of each; Section 2 will, in turn, characterize these modules at a somewhat more detailed level.

For all modules but the first in the sequence (i.e., the emigration module) one enters a given module after having finished all of processing, and the implementation of all of the events, associated with the previous module of the sequence. The modules process the population family by family, operating on the individuals within each family as appropriate.

Internally, most of the modules follow the general structure shown in Figure 1-2 (page 12), and described in more detail below. For reasons of processing efficiency, modules generally have a piece of "screening" code that allows them to skip over families or individuals for whom their events are irrelevant. For example, such code may allow skipping the increment to the age of an individual (just) born in the current simulation year (lest, given the order of the modules, one never have any zero-year-olds, i.e. newborns, in the population pyramid). DYNACAN's family-oriented approach, with its structure of processing individuals within families, is largely responsible for the steps (4) and (5) in the diagram not being treatable as a single step.

Figure 1-2: DYNACAN-B Structure of Typical Probability-Oriented Module

```
                        ┌──────────────────┐
        ┌──────────────→│  From Previous   │
        │               │     Module       │
        │               └──────────────────┘
        │                        │
        │                        ▼
        │   ┏━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┓          ┌──────────────┐
        │   ┃  ┌────────────────────────────┐ ┃          │   Inputs:    │
        │   ┃  │ 1: Data Pass 1: Accumulate │←┃──────────│     Run      │
        │   ┃  │      Expected Events       │ ┃          │ Parameters,  │
        │   ┃  └────────────────────────────┘ ┃          │ Assumptions  │
        │   ┃               │                 ┃          └──────────────┘
        │   ┃               ▼                 ┃
        │   ┃  ┌────────────────────────────┐ ┃
        │   ┃  │ 2: Calculate Expected      │ ┃
        │   ┃  │    Event Rates             │ ┃
        │   ┃  └────────────────────────────┘ ┃
        │   ┃               │                 ┃
        │   ┃               ▼                 ┃          ┌──────────────┐
        │   ┃  ┌────────────────────────────┐ ┃          │  Exogenous   │
        │   ┃  │ 3: Calculate Alignment     │←┃──────────│  Alignment   │
        │   ┃  │    Factors                 │ ┃          │  Parameters  │
        │   ┃  └────────────────────────────┘ ┃          └──────────────┘
        │   ┃               │                 ┃
        │   ┃               ▼                 ┃
        │   ┃  ┌────────────────────────────┐ ┃
        │   ┃  │ 4: Data Pass 2: Implement  │ ┃
        │   ┃  │ Events Using Aligned Probs │ ┃
        │   ┃  └────────────────────────────┘ ┃
        │   ┃               │                 ┃
        │   ┃               ▼                 ┃
        │   ┃  ┌────────────────────────────┐ ┃
        │   ┃  │ 5: Implement Family        │ ┃
        │   ┃  │    Structure Changes       │ ┃
        │   ┃  └────────────────────────────┘ ┃
        │   ┗━━━━━━━━━━━━━━━━│━━━━━━━━━━━━━━━━━┛
        │                   ▼
        │          ┌──────────────────┐
        └──────────│ 6: To Next Module │
                   └──────────────────┘
```

TypModul,CFL

The steps shown in Figure 1-2 carry out the following functions:

1. Data Pass 1: Accumulate Expected Events. DYNACAN-B processes the input database, passing each individual or family through the module. For

families or individuals passing its screening "admission" tests, it uses its equations to calculate preliminary probabilities of "the module's event" occurring for each individual or family. Because alignment factors (described in more detail below) tend to be largely stable from simulation year to simulation year, these preliminary probability calculations use the alignment factors for the module from the previous year. For a typical module dealing with the probabilities of events, it accumulates counts of the possible and expected numbers of events by ***alignment group***. By way of illustration, an alignment group for fertility might consist of all those women in Quebec who are aged 25 through 29 during the simulation year. It also accumulates the numbers of expected events for a grid of small increments to the argument of the equation's logistic representation. In this way, the step conducts a sort of "on the fly sensitivity analysis" that can be used to align the numbers of expected events for each of the alignment groups. However, the step, focussed on data gathering, does not actually implement any of these events, because the alignment remains to be implemented.

2. Calculate Expected Rates. After the first data pass for the module, DYNACAN-B calculates the expected event rates for each of the alignment groups, including, for each group, the expected rates for the grid of adjustments to the function's argument.

3. Calculate Alignment Factors. Still prior to the second pass through the data, for each alignment group, DYNACAN-B uses the grid of expected event rates as compared to the exogenously specified target rates for the event for the group. It derives an optimal adjustment factor that will bring the year's expected rate for the group into line with the group's target event rate. The alignment factor is an additive factor in the logistic function for the equation. This approach guarantees that the outputs that result from the calculation can legitimately be interpreted as probabilities.

4. Data Pass 2: Implement Events Using Aligned Probabilities. DYNACAN processes the input database through the module a second time. In this second pass, the probabilities generated by the module's equations are adjusted, within each of the relevant alignment groups, by the alignment factors calculated in step 3. For each individual and each event, the random variable is compared to the now-aligned event probability to determine if the event occurs. Variance reduction also takes place in this second pass, with the module's event probabilities occasionally tweaked slightly to nudge the actual numbers of events in the direction of the desired, expected numbers of events. It takes only a small amount of nudging to impose a considerable degree of variance reduction.

5. Implement Family Structure Changes. DYNACAN updates the in-memory version of the database to reflect the family structure consequences of any events that affect linkages between individuals' records within families, e.g. births, deaths, marriages, divorces, leaving home, or that affect the numbers of data records for families or individuals within families. Steps 4 (event generation) and 5 (family structure consequences) are carried out, as appropriate for those events that are simulated to occur, for each individual or family in the data base. Between them, they give effect to the events associated with that module's contribution to aging the database by a year.

Then Proceed To the Next Module. At the end of each module, DYNACAN-B has adjusted its "in memory" version of the data base for the population for that year to reflect that module's contribution to the database projection. With the processing of this module complete, unless it is the last one in the sequence (labour-force participation, earnings and weeks worked), DYNACAN-B then executes the next module in the intra-year sequence.

D) Summary Level Descriptions of the Several Modules: The preceding enumeration of DYNACAN-B's modules was sufficient to indicate the overall nature of the modules. However, it didn't convey much about how the modules accomplish their tasks. The descriptions here, though not as detailed as those in Section 2 of this high-level technical documentation, provide a bit more detail. Once again, the descriptions appear in the same (Table 1-1) order in which DYNACAN-B executes them.

☝ **Emigration:** Each simulation year, this module decides for each family in Canada whether the family will be selected as emigrating from Canada during the year. Members of such families emigrate together, *en famille*. Based on government emigration statistics, DYNACAN chooses data base families that look like emigrating families, and ensures that the aggregate numbers of persons, by age, gender and region, reflect its alignment targets.

☝ **Pseudo-immigrants/emigrants:** Given the methodology used by Statistics Canada, and the basis on which immigration statistics are collected, one cannot begin with a known population stock, e.g. from a census, add in the official flows for births, deaths, immigration and emigration, and arrive at the next official stock numbers. If one uses only the official flows, then over time the stocks drift from their known values. DYNACAN-B uses "pseudo-immigrants/emigrants" to reconcile the official population stocks and flows. Though their age-by-gender numbers are small in any one year, over time they make a difference and so have to be included in DYNACAN just as they are in the CPP actuarial valuation model, ACTUCAN. Generally speaking, the

pseudo-immigrants/emigrants reflect population flows such as returning emigrants, and illegal immigration, with the aggregate net flow in a given year generally being positive, i.e. corresponding to net immigration. The algorithms used for this process are effectively identical to those used for immigration and emigration.

✌ **Interprovincial Migration:** Each simulation year, this module decides whether a family residing in Quebec or the rest of Canada will move (i.e. to the rest of Canada from Quebec, or to Quebec from the rest of Canada). The equations for these probabilities take into account a variety of family characteristics. Alignment takes place using age-sex distributions from ACTUCAN.

✌ **Immigration:** Each simulation year, this module decides for each family in Canada whether the family will be marked for cloning to represent a new immigrant family entering Canada. As with emigrants, members of such families immigrate together, *en famille*. Cloned immigrant families generally retain the characteristics of the "donor family," apart from their histories of CPP contributions. However, as in the emigration and interprovincial migration modules, in an adjustment for "chunkiness" in the model, the module may retain the shortfalls or excesses of immigrants with respect to their targets, carrying them forward to the next simulation year.

✌ **Mortality:** Each simulation year, this module decides whether an individual living at the start of the year will be selected to die during the year. The probabilities of death depend on calendar year, age, gender, marital status, income, disability status, and region. We align the mortality rates implicit in the mortality equations using CPP Actuarial Valuation Model (ACTUCAN) rates that depend primarily on age, gender and region, with some special considerations for disability status.

✌ **Fertility:** Each simulation year, this module decides for each woman whether she will bear a child during the year. The probabilities depend on the woman's age, region, marital status and family structure, educational status, and labour force status. We align the fertility rates implicit in the equations to those for ACTUCAN, within subgroups defined in terms of age group, region and simulation year.

✌ **Aging:** Each simulation year, each individual who has not died or just been born, becomes a year older. No alignment is needed.

✌ **Marriage and Remarriage:** This module has two sub-components. First, each simulation year, it identifies appropriate numbers of persons, separately

by gender, as being "in the marriage market." The probability of entering the pool depends on a person's characteristics, e.g. age, education, earnings level, marital status (never-married, separated, divorced, widow(er)ed) and the presence of dependent children. Second, using a compatibility index based on the individuals' characteristics, the module serves as "matchmaker" to create specific couples from the gender-specific pools of individuals in the market that simulation year. We align marriage and remarriage rates using Statistics Canada vital statistics information on the age and gender-specific fractions of people marrying in a given year, and some U.S. data from CORSIM where Canadian data are not available.

☞ **Separation and Divorce:** Each simulation year, this module decides for each couple whether that couple will separate/divorce during the year. The probabilities of splitting depend on the characteristics of the couple, with explicit provision for the length of the union. Alignment for this module uses data from U.S. vital statistics where no comparable Canadian data are available.

☞ **Leaving Home:** Generally speaking, children will eventually leave the family home, founding new households either as unattached individuals or as members of a newly formed couple. Each simulation year, this module looks at children living at home and decides whether they will leave that home. There is no alignment for this module. The current version of DYNACAN does not allow children to later return to the home.

☞ **Education:** Each year, individuals currently in school may complete another year of school or leave school, potentially to join the labour force. Children enter school at age 6. This module decides, for each individual already in school at the beginning of the year whether to increment that person's years of schooling by one additional year during the simulation year. DYNACAN assumes a specific, fixed relationship between the years of education completed and educational attainment (e.g. a high school diploma is deemed to be completed once an individual has completed twelve years of education). In addition, the module provides for an "instant upgrade" for selected 32-year-old individuals to reflect the impact of later-life education. The DYNACAN team has not yet implemented alignment for the education module.

☞ **Disability:** Based on statistics for Canada Pension Plan disability benefit emergences, this module selects and marks individuals deemed to have become disabled by the standards of the Canada Pension Plan. The module also marks disabled individuals deemed to have recovered. However, the actual calculations of CPP disability benefit eligibilities, and of the benefit levels where relevant, take place in DYNACAN-C.

☞ **Exclusion of Certain Females from the Labour Force:** This module is DYNACAN-B's dynamic counterpart to DYNACAN-A's labeling of certain women as excluded from entering the labour force. Each simulation year, it marks a proportion of those women turning 21 that year as being excluded from future labour force participation. These women are selected based on marital status, the presence of children, and labour force participation to that point in their lives. The proportions are chosen so that, in future simulation years falling in the historical period, an appropriate fraction of women at various ages will "qualify" as having never worked (relative to extant survey data). The proportion of the population so marked gradually declines over time, as is consistent with increasing labour force participation by women.

☞ **Retirement:** Once individuals reach age 60, each year some of those formerly working retire for CPP purposes, and begin drawing CPP retirement pensions. Each simulation year, this module decides, for each individual 60 or older and not yet retired for CPP purposes, whether that individual will retire, for CPP retirement benefit purposes, during the year. This module ensures that the proportions of persons retiring are consistent with those assumed by the CPP Actuarial Valuation Model, ACTUCAN, for age, gender, current retirement status, and simulation year, including ACTUCAN's assumption that individuals retire for CPP purposes by age 65. It includes the special provision that persons in receipt of a disability pension are assumed always to convert from a disability pension to a retirement pension at age 65 if they are still alive at that point.

☞ **Employment and Earnings:** Each simulation year, this module decides, for each individual, whether or not the individual will work, how many weeks the individual will work during the year, and what the earnings from that work will be. It is the single most complex of DYNACAN-B's modules. The simulated weeks-worked and earnings outputs depend on several variables, from the person's age, gender, education and family structure, through retirement status, disability status, and previous earnings history. We align the calendar year and gender-and-age-specific employment earnings distributions to data from ACTUCAN distributions drawn from administrative Record of Earnings data. Further, longitudinally, the module should generate lifetime earnings scenarios that are individually and collectively realistic, so that they serve as valid inputs for the subsequent calculation of CPP contributions and benefits, and for the CPP's general drop-out. The use of previous labour-force attachment in the participation decision helps to ensure the proper degree of labour-force stickiness over time.

In addition to producing annual representative cross-sectional population databases, DYNACAN-B can produce "flat files" containing user-specified pieces of information

about individuals or families. These flat files have one line/record for each relevant person or family for each simulation year, and are ordered by person (or family) identifier. One particularly important flat file produced at the level of the individual gives each person's age, sex, earnings, disability status and province of residence. As described below, this particular file plays a crucial role in identifying those individuals who must make CPP contributions, and who may receive CPP benefits.

DYNACAN-C (**C**alculate CPP Contributions and Benefits): DYNACAN-C typically uses as its input the series of population databases output by DYNACAN-B. (The much less typical usage sees DYNACAN-B and DYNACAN-C executed each year, in an alternating sequence, in order to permit feedback between the CPP and the model's demographics.) Proceeding from the events and characteristics for the populations of individuals, DYNACAN-C calculates, simulation year by simulation year, individual by individual, all of the contributions payable to the CPP, and all of the benefits paid by the CPP to the individuals. Central to this module is the design criterion that seeks to keep the contribution/benefit calculations as close as possible to those actually used in the administration of the CPP. Suitable collation of these representative contributions and benefits values, in conjunction with the characteristics of the individuals to which they apply (as they age through time), then yields virtually any desired array of cross-sectional and/or longitudinal results desired by the client.

Despite its singular name, DYNACAN-C is in fact a composite of several different processes executed in sequence. The major component processes, presented here in the order of their execution, serve the following functions:

☞   **exclqueb**: This program processes the information in the DYNACAN-B flat file (noted above) that contains the regional information (Québec vs. Outside Québec status) for individuals in the simulation. It identifies all of the persons who are required to make CPP contributions on their earnings in certain years, as well as all of those who are eligible to receive CPP retirement benefits (because they made their last CPP/QPP contribution outside Québec). It indicates the first and last years for which the individual is relevant for either CPP contributions or benefits. The output from this process is ordered by personal identifier, as is required for processing by subsequent DYNACAN-C components.

☞   **cpp**: This program is the heart of DYNACAN-C. It reads the DYNACAN-B output, and calculates all of the relevant contributions and benefits for those persons. (Note that the some of the population will be relevant for the QPP rather than the CPP, and some will not be relevant for either.) The details are far too voluminous to include in this high level documentation, but the calculations (for the current CPP system) mirror as closely as possible the same ones used by Income Security Programs Branch in its administration of actual CPP benefits. Section 4 of this high-level technical documentation provides substantial additional detail. The output from this **cpp** program takes the form of a set of

files, each up to 1.5Gb in size; for most runs a single file is sufficient to hold all of the relevant information. There is one record for each individual; it holds, for all years of the simulation, the individual's contributions, and all of the benefits, by benefit type, received by the individual.
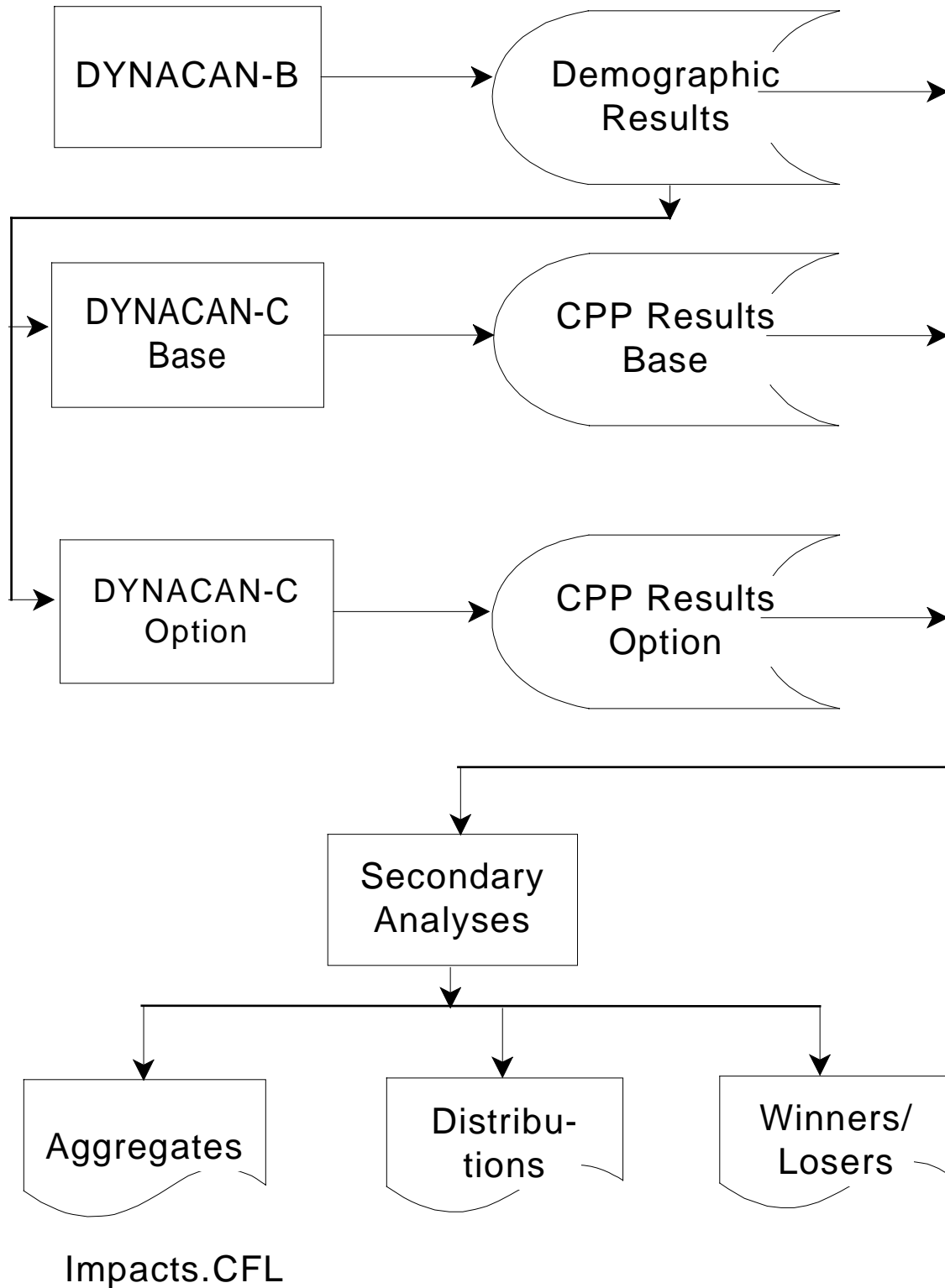
☝ **postproc**: This program massages the disk-cache files produced by **cpp** to discard the irrelevant portions (e.g. excluding the data for individuals not identified by **exclqueb** as relevant for the CPP). It generates text data file containing each person's unadjusted pensionable earnings, contributions, benefits by type, age, sex and personal identification number, as well as the simulation year for which that information is relevant. The information is normally ordered by person identifier, and within persons by simulation year.

☝ With the calculation of the CPP contributions and benefits complete and the results written to a well-structured file, DYNACAN-C's work is essentially complete. For a typical analysis, there remains the task of ascertaining the particular tables required by the client, and then extracting the relevant information from the **postproc** output and/or the DYNACAN-B output files, and collating it into the tabular or graphic formats suited to the clients' needs. These considerations are also relevant for comparative policy analyses using DYNACAN. Although the open-endedness of this stage prevents any comprehensive treatment in this high level documentation, Section 5 provides some illustrative detail.

We have not addressed here that there are multiple configurations of DYNACAN-C. They exist to respond to efficiency and resource management issues associated with, for example, running DYNACAN-C in a lower RAM environment or in a feedback mode with DYNACAN-B. Additional modules are relevant for some of these modes, but their identification and description is not appropriate for a high level documentation such as this.

With the descriptions of the DYNACAN components complete at this level, it is useful to examine briefly how the DYNACAN-B and DYNACAN-C components work together in a typical policy impact analysis. Figure 1-3 (page 21) displays the key relationships. Except when demographic feedback is required, a single run of DYNACAN-B produces, from the starting database, the common demographics and earnings data for the base and option systems. (The use of feedback and its implications are too complex to allow its inclusion in this high level documentation.) Parallel runs of DYNACAN-C calculate the CPP contributions and benefits for those two systems. Secondary analyses, potentially drawing on several kinds of output files and software, assemble the specific outputs needed for the policy option being examined.

Figure 1-3: Schematic for Typical
DYNACAN Impact Analysis

```
┌──────────────┐          ⟩ Demographic ⟩
│  DYNACAN-B   │ ────────▶   Results
└──────────────┘          ⟩            ⟩ ────────▶
                                  │
                                  ▼
┌──────────────┐          ⟩ CPP Results ⟩
│  DYNACAN-C   │ ────────▶    Base
│    Base      │          ⟩            ⟩ ────────▶
└──────────────┘

┌──────────────┐          ⟩ CPP Results ⟩
│  DYNACAN-C   │ ────────▶   Option
│    Option    │          ⟩            ⟩ ────────▶
└──────────────┘

                    ┌──────────────┐
                    │  Secondary   │
                    │  Analyses    │
                    └──────────────┘
              ┌──────────┼──────────┐
              ▼          ▼          ▼
        ⟨Aggregates⟩ ⟨Distribu-⟩ ⟨Winners/⟩
                     ⟨ tions  ⟩ ⟨ Losers ⟩
```

Impacts.CFL

Other Key Model Characteristics: Several of the issues implicit in the preceding discussion of DYNACAN's components warrant some additional precision, even here in the high level documentation. However, Section 5 provides considerably greater detail than it is appropriate to include here.

☞   *Monte Carlo Nature of DYNACAN Simulations*: As noted above, DYNACAN uses Monte Carlo simulation in generating its events. Pseudo-random deviates determine whether a given event (e.g. death) will happen for a given individual in a given simulation year. If the internal deviate generator were to be seeded with a different starting value, different events would occur, and slightly different aggregate and distributional outputs would result. The variation occurs because, for example, individuals might die at different points in time, marry different partners, have different earnings or different numbers of children, etc. This variation means that, even though individual DYNACAN runs made from a common starting seed value are generally replicable, the outputs from any one run are properly seen as the result of one particular draw from a distribution of possible outcomes. The "means and variances" of those distributions of outputs are, however, are not known beforehand. Further, the computational costs of actually making significant numbers of repeated draws have been seen as generally prohibitive. This characteristic variability of DYNACAN results stands in sharp contrast to the outputs from deterministic models such as the CPP Actuarial Valuation Model, ACTUCAN. There, any set of runs using common assumptions and parameters will generate precisely the same resulting outputs. However, since the costs of calculation are falling steadily, it is becoming more conceivable that one might average the results of multiple, equivalent, runs varying only in the random number seed, thereby reducing the degree of inherent variability.

Alignment to the CPP Actuarial Model: Also as noted above, a consistency of results with those from the ACTUCAN (when one uses equivalent assumptions) is a particularly important design criterion for DYNACAN. Thus, for many of DYNACAN-B's modules we align the expected rates for DYNACAN events so that they closely match external target rates. Most of those alignment targets come from the rates used in ACTUCAN, or from the assumptions used for the run. In a representative application of alignment, one compares the expected rates for events that will occur in DYNACAN to the corresponding rates in ACTUCAN. Based on this comparison, one then adjusts the DYNACAN probabilities via an increment to the logistic function of the probability equation for the event. Recall that the quantitative value of the increment is chosen to make the expected and target rates for the event equal across the alignment group. The nature of this adjustment was described above in the section dealing with the structure of a

typical DYNACAN-B module, and will not be repeated here. Section 2 of this document provides additional detail on this form of probability alignment.

Two other forms of alignment are also relevant. The first of these, also noted above, involves earnings. In a given simulation year, the target distribution of earnings for earners grouped by age and gender is given by ACTUCAN, as derived from administrative earnings data that have proved surprisingly stable over time. The earnings alignment uses the inverse cumulative density function for this target distribution. For example, suppose an individual's earnings as generated by DYNACAN's equations place him one-third of the way through the DYNACAN distribution of earners ordered by level of earnings. Then the earnings alignment algorithm assigns him the level of earnings appropriate to someone one-third of the way through the target earnings distribution.

The remaining form of alignment is relevant for immigration and emigration. Under it, candidate families for emigration, or for cloning to simulate immigrants, are over-sampled in the data to produce a pool of possibilities. However, the alignment targets for a year correspond to specific age and gender combinations. The alignment algorithm chooses families from the pools of candidates so as to approximately meet the desired distributions by age and gender. Any mismatches are carried forward to affect the next year's targets, so that errors do not cumulate.

☞ *Variance Reduction*: As discussed above, the use of Monte Carlo simulation means that the output from a run of the model should be properly interpreted as that of one particular draw from an unknown distribution. Under the present hardware configuration, it would be costly in time and computer resources to run the simulation many times to ascertain some average result. Consequently, one would like to alter the logic of the model so that the various outputs, though to a certain extent still random, behave as if they came from a distribution tightly centered around that unknown central tendency. Most of DYNACAN-B's probabilistic components use variance reduction techniques to ensure that the actual numbers of events closely approximate the expected numbers, while still retaining the positive features of Monte Carlo simulation. As described elsewhere in this documentation, the variance reduction techniques have two major characteristics, both of which depend on the model's keeping running track of the numbers of events that have occurred and the numbers expected to that point.

- First, the variance reduction algorithm may slightly adjust the probabilities of individual events if that adjustment would bring the actual number of generated events closer to the expected number.

- Second, the model implements a form of "institutional memory" so that excesses and shortfalls for a particular class of events in one period will affect the target probabilities for numbers of related events in the next

period. E.g. if too few 80-84-year-olds die in one year, the model slightly increases the probability of 80-84-year-olds dying during the next simulation year.

The combination of these two variance reduction techniques results in a substantial tightening of the dispersion of outputs.

☞ *Comparative Nature of Most Policy Runs*: It is the essence of most policy analysis that one wishes to ascertain the impacts of changing some aspect of the existing system. One may measure these impacts in a variety of ways, including numbers of winners/losers, dollar impacts or shifts in key distributions. In the present configuration of DYNACAN this impact assessment typically corresponds, as in Figure 1-3, to running two different versions of DYNACAN-C (with one run for the existing system, and another run for a proposed alternative/option system). Both runs use the same set of DYNACAN-B population databases with its inherent demographic and economic evolution. Separate filtering programs then compare individuals' CPP contributions and benefits in one DYNACAN-C output with the corresponding contributions and benefits in the other to ascertain the impacts at the level of the individual. Appropriate aggregation/collation then yields the relevant total, distributional and longitudinal impacts at the desired level of unit of analysis.

1.3   Project Context

There is considerable respect at HRDC and elsewhere for the CPP Actuarial Valuation Model, ACTUCAN, and for the accomplishments of the Office of the Chief Actuary, which has regularly and successfully used it in support of CPP policy analyses. The DYNACAN team emphatically shares this respect. Impressively, these successes have occurred despite the CPP Actuarial Valuation Model's having been constructed to address actuarial needs rather than policy needs. However, it was recognized that the CPP Actuarial Valuation Model often did not offer sufficient detail for a variety of policy-critical questions. Moreover, the cell-based nature of ACTUCAN precluded enhancements to enable it to address those policy-critical questions.

The development of DYNACAN was essentially demand driven. There were important policy issues that could not be effectively addressed in the desired level of detail by existing models such as the CPP Actuarial Valuation Model, nor by natural extensions of those existing models. Specific HRDC and OSFI requirements included:

!   Addressing interactions of the CPP with the full system of tax/transfer programs, including income taxes, income-tested benefits and other programs, at multiple levels of government.

! Performing winners/losers analyses. Especially in the current environment of cutbacks and targeting, it is necessary to be able to identify the numbers and characteristics of the winners and the losers, and the sizes of their gains or losses.

! Identifying cross-subsidies for a variety of groups, and across levels of government, specifically including cross-cohort, and cross-time subsidies that reflect the overall tax burden from all sources of taxation. This identification includes group-specific impacts that arise from income-tested benefits, as well as from direct income taxes and consumption taxes. It may involve individuals' lifetime tax burdens, as well as the taxes and clawbacks occurring within a given year.

! Providing information in customized formats that the policy-developer clients require. Given the varying foci of policy initiatives, these results needed to be flexible, inexpensive and easily exportable in machine-readable formats.

! Answering 'what-if?' questions (usually with a strong emphasis on comparing an option system to the existing CPP).

! Presenting a sufficient level of detail for analyses, one that goes beyond the necessarily limited numbers of categorical variables that cell-based models are able to manage.

! Ensuring consistency with established data and models, and particularly with OSFI's CPP Actuarial Valuation Model, so that the government can speak on CPP policy issues with a single voice.

After careful consideration and exploration of all possible options, HRDC and OSFI reached the conclusion that longitudinal dynamic microsimulation is the only technology that is able to address all of the issues noted above. The costs for building, maintaining and using such a model, while significant, are minuscule compared to the cost of even a single bad CPP policy decision based on inadequate analysis.

Although DYNACAN is strongly CPP-focussed, it is expected there will be significant, positive policy-analytic benefits to other areas, though of course the funding for non-CPP applications will need to come from sources other than the CPP.

Abbreviated Project History

Given the policy requirements identified above, it was recognized that existing policy tools and models were not sufficient to address the policy issues in sufficient detail. Investigations suggested the feasibility of constructing a suitable new longitudinal dynamic microsimulation model, one now known as DYNACAN.

In 1990, prior to making a decision to either build a new model from scratch or to adopt an existing model as a template, or to make any significant expenditures, the Income Security Programs Branch (of what is now HRDC) carried out a careful review and an initial needs-determination. The necessity for the active involvement of OSFI became immediately apparent. Subsequently, an external consultant, contracted to provide an independent, objective perspective, confirmed the need for the model. That consultant offered the tentative conclusion that construction of a suitable model was feasible, and assisted in the development of preliminary cost estimates. Proceeding from this conclusion, OSFI and HRDC conducted a formal feasibility study that confirmed both the project's feasibility and the soundness of the cost estimates.

OSFI and HRDC project management made the strategic decision to construct a prototype model as a feasibility demonstration, and to use an existing model as a template for that prototype. The prototype would initially simulate the CPP independently of other tax/transfer programs, but would be sufficiently robust to serve as a foundation for subsequently simulating the interaction of the CPP with those other programs. Models from three countries were investigated as to their suitability and adaptability to HRDC's template requirements.

By 1995, the initial staffing was complete and an existing operational model, CORSIM, developed under the leadership of Professor Steven Caldwell of Cornell University, was acquired to serve as the template for the prototype. The project team chose the name DYNACAN, and began the Canadianization of the model's demographics, labour force participation and employment earning modules, as well as the creation of new modules for immigration/emigration.

During 1996, Canadianization of the demographics continued, and the construction of the CPP module began, along with an initial validation of the model. DYNACAN was able to contribute marginally to the 1996 quinquennial review of the CPP. The remainder of 1996 saw an aggressive effort to ensure the critical design goal of consistency with the CPP Actuarial Valuation Model.

Those consistency oriented efforts continued in 1997. Also addressed in 1997 were a major restructuring of the model's calculations, termed "remodularization" in which all of the DYNACAN-B modules were rewritten so that all individuals/families are processed through each successive module. Previously, families passed, one at a time through all of

the modules before the next family was processed. The remodularization not only provided significant improvements in computational efficiency, but resulted in code that was shorter, more standardized and more maintainable. Moreover, as anticipated, the remodularization improved the effectiveness of the several alignment procedures.

The year 1998 saw a major improvement to the education module, completion of the remodularization, and major reworkings of the immigration/emigration and earnings modules. Substantial effort was also devoted to validation, including longitudinal earnings validation, benefit take-up, and a reworking of the DYNACAN-C CPP calculations in preparation for DYNACAN's applications to the "Track 2" CPP options. The last half of the year saw a planned transfer of the DYNACAN project to the Strategic Policy Branch at HRDC.

Organizational Partners and Project Funding

There have been three major organizations involved in the development of DYNACAN:

1)   During the initial development of DYNACAN, the DYNACAN project team operated under the direction of the Chief Actuary in the Office of the Superintendent of Financial Institutions. This placement was considered appropriate in view of the high priority placed on consistency of outputs with those from the existing CPP Actuarial Model.

2)   The Strategic Policy Branch of HRDC is heavily involved as the primary client for DYNACAN results, and as an experienced practitioner in the development, maintenance and application of policy models.

3)   The Income Security Programs Branch (ISP) at HRDC is responsible for the administration of the Canada Pension Plan; it first initiated the DYNACAN project, and provides significant portions of the CPP-related data that are used for model calibration and validation.

Consistent with DYNACAN's strong CPP focus as regards its outputs, and their usage, DYNACAN's funding comes from the Canada Pension Plan. Total development costs are expected to reach about 6 million 1992 dollars, spread over about a five year development schedule. Accountability has been ensured via the CPP's standard procedures relating to OSFI's recovery of the costs it incurs in the course of providing services to the CPP. These are governed by a formal memorandum of understanding with respect to the services provided and the recovery of costs for those services. As well, DYNACAN's annual budget projections for the development period were reviewed and approved in advance by a Project Board. That board, the primary governance body for DYNACAN, included representatives from the OSFI (the builder), Strategic Policy HRDC (the major client) and Income Security Programs HRDC (in its capacity as both

data supplier and client), with the latter's representation also including the CPP Comptroller.

1.4 Role of ACTUCAN (CPP Actuarial Model) with respect to DYNACAN

As will have been apparent from the repeated references to the CPP Actuarial Valuation Model, ACTUCAN, that model and its outputs are intimately linked to DYNACAN. OSFI's 1995-96 Annual Report for the Valuation Unit and the Social Policy Unit indicates the basic nature of the relationship: "The Social Policy Unit continued its involvement in a five-year project, sponsored by the Department of Human Resources Development, for the development of a micro-simulation model which determines the financial effect on individuals of CPP policy proposals once analyzed in aggregate by the Valuation Unit." It is useful to conclude this introductory section by characterizing more explicitly the nature of that relationship.

☞ The CPP Actuarial Valuation Model is an actuarially oriented, cell-based model constructed and maintained by the Office of the Chief Actuary at OSFI to support actuarial CPP analyses that are required by the CPP legislation. Appendix B of periodic statutory actuarial reports, e.g. the 15th report on the status of the Plan as of December 31, 1993, provides an extensive description of the model's logic in terms of its data, assumptions and methodology. There is an extensive, on-going program of validation to ensure that ACTUCAN produces results consistent with administrative program and demographic/economic statistics. External reviews of the CPP Actuarial Model indicate that it competently satisfies the objectives for which it was constructed.

As well, beyond using the model for strictly actuarial purposes, the Office of the Chief Actuary has shown considerable diligence and ingenuity in harnessing ACTUCAN to address policy needs where the relevant calculations fall within its scope. Nevertheless, ACTUCAN has relatively little to say about important policy issues such as family structures, winners/losers, or the interaction of the CPP with other tax/transfer programs. We developed DYNACAN to fill these needs.

☞ DYNACAN has been designed to produce results consistent with those of the CPP Actuarial Valuation Model when equivalent assumptions are used. Toward this end, several of the DYNACAN modules are constructed to use parametric assumptions of the same form as those used in the CPP Actuarial Model. For example, ACTUCAN's parametric specification of birth rates by age range and simulation year also appear in DYNACAN, though DYNACAN's fertility equations provide for a much richer and more faithful characterization of the particular mothers to whom the children are born.

✋ In the most aggressive instance of attempting to ensure the consistency of results between the two models, event rates for several of the key demographic and employment earnings flows in DYNACAN are aligned to exogenous target rates established by the CPP Actuarial Valuation Model. In essence, DYNACAN's equations provide the dispersion for the simulated events, but DYNACAN chooses to adopt ACTUCAN's assumed aggregate rates for the events. The major instances in which ACTUCAN's rates are normally used to align DYNACAN's event probabilities and earnings include --

- mortality rates by age, gender, region and year,
- fertility rates by age group, region and year,
- migration rates by age, gender, region and year,
- earnings distributions by age, gender, region and year, and
- retirement rates, by age, gender and year.

✋ ACTUCAN provides many of the standards for validating DYNACAN's performance in projecting aggregates, though, in turn, many of those standards for ACTUCAN derive from survey and administrative data. Illustratively, a quite non-exhaustive list of such standards includes --
- population by age/gender/year,
- births, deaths and net immigration by age/gender/year,
- employment earnings distributions by age/gender/year
- numbers of contributors and beneficiaries by age/gender/year,
- amounts of contributions or benefits by age/gender/year , and
- pay-as-you-go rates for the CPP as a whole and for individual types of benefits.

The decision to rely on the CPP Actuarial Valuation Model's results as the standard for such validation measures is founded on
- the extensive validation that ACTUCAN undergoes against a wide range of historical administrative and survey data,
- the consistency of its calculations with standard actuarial practice, combined with
- a confidence that arises from its methods having repeatedly met the theoretical test of external review, and
- the practical test of ongoing comparisons of its projected results with subsequent administrative statistics.

✋ The CPP Actuarial Valuation Model efficiently projects the changes in required CPP contribution rates as they flow from prospective changes to the CPP. At the present time, DYNACAN contains no mechanism for replicating such altered contribution rate schedules. Instead, DYNACAN takes as inputs the contribution rate schedules derived by ACTUCAN. Consequently, DYNACAN is presently best suited for providing distributive and policy-oriented detail for policy options

that have already been explored by ACTUCAN with respect to their broader financial impacts and contribution rate implications.

☞   The CPP Actuarial Valuation Model also includes a component for the projection of the CPP fund and the portfolio of bonds in which it is invested. As well, ACTUCAN provides the basis for the projection of the unfunded liability that the CPP would exhibit, were it to operate as a conventional fully-funded pension plan. DYNACAN, as it presently stands, includes no such capacities, nor is any replication of those capacities currently planned.